Sim, Yiing Luen (2005) A method for treating dependencies between variables in a simulation risk analysis model. PhD thesis, Middlesex University. [Thesis]

Final accepted version (with author's formatting)

A Method for Treating Dependencies Between Variables in a Simulation Risk Analysis Model

A thesis submitted to Middlesex University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

By

Yiing Luen Sim
Middlesex University Business School

December 2005

Abstract

This thesis explores the need to recognise and represent accurately the interdependencies between uncertain quantitative components in a simulation model. Therefore, helping to fill the gap between acknowledging the importance of modelling correlation and the actual specification and implementation of a procedure for modelling accurate measures of Pearson's correlation became the main aim of this research.

Two principal objectives are stated for the developed Research Correlation Model ("RCM"): (1) it is to generate Pearson-correlated paired samples of two continuous variables for which the sample correlation is a good approximation to the target correlation; and (2) the sampled values of the two individual variables must have very accurate means and variances.

The research results conclude that the samples from the four chosen distributions that have been generated by the RCM have highly acceptable levels of precision when tested using χ^2 tests and others. The results also show that an average improvement in precision of correlation modelling was over 96 percent. Even with samples as small as 10 the worst case correction factor is only just less than 90 percent, with the average correction factor being over 96 percent overall, so that the contribution made by the RCM here is quite impressive.

Overall the analysis shows that in the case when the sample size is 10, the RCM consistently generates samples whose correlation is so much more precise than that generated by @RISK. The smallest of all the observed ratios of improvements of the RCM in comparison with the use of @RISK is 2.3:1, in just one case when the medians were being compared. The average improvement ratio exceeded 100.

It is concluded that the aim of specifying, formulating and developing a Pearson correlation model between a pair of continuous variables which can be incorporated into simulation models of complex applications has been achieved successfully.

Table of Contents

List of Figures

List of Tables

List of Appendices

Acknowledgements

I am grateful to people who have helped in various ways in my study. First of all, it has to be Dr. Alan Pryor, my Director of Studies, who has the credit for taking me diligently on my journey through this research project. I am indebted to him for his patience, kindness, guidance, support and advice, which provided a never-failing source of encouragement and stimulus. I also want to thank David Jarrett, my second supervisor, and Professor Jonathan Liu, for their helpful and stimulating suggestions.

Most of all, I must thank my fiancé, Tony Huang, who inspired and supported me one way or another during my research and also showed great understanding and tolerance in the face of my increased obsession with the research project. I would also like to thank my parents for giving me full emotional support while I was in the down side of my research journey.

My thanks also go to Martin Kender and Frida Attram for their administration support. Thanks would also go to anonymous people who somehow provided assistance and suggestions whenever I needed them. I hope they all find it was worth the effort.

Chapter 1: The Thesis

1.1 Introduction

This thesis explores the need to recognise and represent accurately the interdependencies between uncertain components in a model of a real world application, whether that model is analytical or so complex that it can only be modelled via some form of simulation.

Uncertain components are variables that are used in a model and which have values that are not certain. I.e. Not deterministic. The level of uncertainty could be affected by internal factors, e.g. change of capital structure, or external factors, e.g. political or economic change.

Quantitative risk analysis (QRA) using simulation is a powerful method for portraying the uncertainty and variability of a problem and for giving one a realistic appreciation of the problem's total uncertainty. One application area is that of capital budgetting, and it will be discussed in section 2.2. The role of capital budgetting in this thesis will be to illustrate or interpret many of the mathematical, statistical, economic, or operational research tools which will be documented within the research.

One of the golden rules of QRA using simulation is that each scenario identified as an output by the simulation must be potentially observable in real life (Vose, 2000).

To illustrate this, suppose there is a model relating interest rates and mortgage rates. They should be positively correlated, so that an outcome of the simulation which allows, for example, a higher interest rate and lower mortgage rate is far from desirable. Thus, disregarding the correlation, which is a measure of interdependencies and the association between these two variables, may result in impossible combinations being generated, in which case the cash-flows calculated would not be practical. As a result, the whole process could waste resources e.g. money, time and effort.

Any simulation model, therefore, must be restricted to prevent it from producing, in any iteration, a scenario that could not sensibly occur.

During the early stages of this research it became very clear that commercial simulation or risk analysis packages, such as @RISK and Crystal Ball, enable some degree of correlation modelling to take place. However, the correlation modelling is based on Spearman's rank correlation technique (Curwin and Slater, 2002), which is appropriate for variables represented on an ordinal scale. Where variables are continuous (such as \$ to £ exchange rates and inflation rates) it would be rather more appropriate to measure correlation using Pearson's product moment correlation coefficient. These simulation models enable accurate representations of the probability distributions of the individual variables (i.e. the marginal distributions) to be generated, and research was carried out to identify any implementations of Pearson's method in commercial simulation or risk analysis packages. Initially nothing relevant was discovered, but eventually the work on NORTA and its derivatives (Cario and Nelson, 1997) was found.

This is discussed later in Chapter 2, but the conclusion drawn there shows that the NORTA approach is very complex, requires the use of a great deal of computing and time-related resources, and is all too often non-robust in the sense that it fails to converge to an acceptable result. This thesis will demonstrate that it is indeed possible to generate multivariate samples whose marginal distributions for the individual variables are specified, such that the simulated sample values fit precisely to their marginal distributions, and for which the achieved product-moment correlation is very accurate.

Indeed, Wall (1997) claims that literature on presenting Monte Carlo simulation often overplays the importance of the choice of which distribution to use to represent input variables which are believed to be uncertain and underplays the importance of assessing and including correlations between these inputs (further illustrations are discussed in Chapters 2 and 3).

This claim by Wall and the desire to seek ways of modelling Pearson's correlation coefficients therefore have become the starting point for this research. Hence, helping to fill the gap between acknowledging the importance of modelling correlation and the actual specification and implementation of a procedure for modelling accurate measures of Pearson's correlation has turned into the main aim of this research.

When this research originally began it was for a short time directed at examining new or better usage of economic parameters within capital budgetting. Rapidly this evolved into an intention to develop an economic risk analysis product suitable for the evaluation of a project such as a proposed new oil field in the petroleum industry, so that the drive would be on the application area, with new or revised 'building block' techniques being sought as appropriate on the way.

Quite soon after this stage, however, the current shortcomings relating to simulating product-moment correlations became apparent, and so the current emphasis emerged: to seek a more accurate and robust and/or less complex means of simulating product-moment correlations, with illustrative examples being drawn from numerous application areas, but principally those of capital budgetting and oil field economic analysis.

1.2 Research question, aim and objectives

1.2.1. The research question

The main research question is:

"How can the relationship between continuous variables be integrated in a simulation model using Pearson's product moment correlation?"

Simulation is a technique of Operational Research which involves using a computer to imitate (simulate) the operation of an entire process or system by randomly generating and recording the occurrences of the various events that drive the system, just as if it were physically operating. This is explained in greater detail in Chapter 2. QRA models using simulation are more complex than the deterministic models that they build on. A major reason for this increase in complexity is that a simulation model is dynamic.

There are a potentially infinite number of possible combinations of scenarios that can be generated by a risk analysis model. However the output from the simulation is only useful if these scenarios are viable in the real situation.

Therefore it is important that some sort of control can be imposed upon the sample values used in each scenario. This will prevent impossible combinations being used in producing a scenario which in return might then be used in decision making.

1.2.2. The research aim

Initially the research is intended to model the correlation between only a pair of continuous variables, so that the above research question leads to the primary aim of this research as follows:

"To specify, formulate and develop a Pearson product moment correlation model between a pair of continuous variables which can be incorporated into simulation models of complex applications."

This model will be known as the 'Research Correlation Model' or RCM. In simulation parlance this is known as an 'Input Model', and contrasts with the bulk of the modelling complexity which goes into the building and specification of the 'Logical Model', for example to generate post-run analyses. It will become clear within this thesis that the general view is that commercial simulation packages are usually helpful and supportive to the user when designing, implementing and testing the logical model, for example with the provision of report writer routines, but in contrast the facilities for input models are either too limiting and elementary, or require lengthy and in-depth input by the user into complex input routines. Thus any advance which reduces the 'black box' element of the input model, or lessens the need for the user to oversee repetitive and high level mathematical and statistical computations is indeed laudable.

When the desirability of accounting for the interdependencies between two variables in a simulation model is recognised, the need for a model such as the RCM that could fulfil the requirement is obvious.

A model by definition is a representation of real objects or situation (Hillier and Lieberman, 2005). Specifying a model includes:

1. Finding a way of expressing the understanding of situations through the use of simplified constructions, the use of language, the use of diagrams or the use of mathematics.

2. Constructing a transformation where outcomes are explained by a range of inputs and assumptions. Assumptions are things that are believed to be true for the model and they are imposed to limit the scope for formulating and developing of a model.

3. Identifying which variables should usefully be modelled. A balance is needed between those inputs that are significant and those that may have some minor effect but do not significantly impact on the problem characteristics.

4. Creating an understanding of relationships between the outcomes from a model with the inputs and assumptions that are affecting the outcomes. These relationships need to be specified in terms of being ‘fit for purpose’ rather than perfectly correct.

5. Establishing the testing and verifying processes that best suit the aim and objective of the RCM. This will include the categories of data that are required to prove the validity of the model.

The actual formulation is best described as the prototype that forms the foundation theoretical approach of the RCM identified in the research aim by combining certain methodologies that will be reviewed and proved to be useful in inventing a distinctive way of generating a pair of correlated sample values.

Later in the thesis, in Chapter 6, the output from the formulation is proven to be acceptable by comparing with the output from a widely used commercial simulation package. The prototype is then extended to cope with broader or more complex modelling problems. This is when the prototype is developed as a computer model.

1.2.3. The principal research objectives

The research aim is to develop a product moment correlation model for two continuous variables, the functionality of which must attain two objectives, as follows:

Research objective 1:

The correlation model must generate samples of pairs of values of continuous variables whose Pearson correlation coefficient has acceptable precision

The correlation model should be able to produce numbers of pairs of correlated sample values, depending on the given type of probability distribution assigned to each variable and the relationship between them in terms of their Pearson correlation coefficient. At the end of the output, the two sets of the sample values generated from the model must have measures of inter-dependencies that are as close as possible to the required correlation coefficient.

Research objective 2:

The correlation model must include a good representation of the uncertain variables

The sample values generated from the correlation model not only need to correspond to the input correlation coefficient but also to abide by the descriptive statistics e.g. measures of central tendency and spread, etc. For example, the calculated sample mean of either of the variables needs to be acceptably close on some scale to the expected value of that variable. Thus, these descriptive statistics are calculated based on the parameters given to the assigned probability distributions.

Being able to show that the output from the model will meet the requirement of the relevant descriptive statistics will ensure that these sample values are truly representative of the input variables. It is therefore an important process to validate if the research invention has been properly done and has achieved its aim.

1.2.4. Supplementary research objectives

Alongside achieving the aim and objectives of this research, there are other objectives to be attained. They are:

- Defining the terminology used in QRA, such as uncertainty, variability and risk.

- Presenting, comparing and contrasting different approaches used in quantifying uncertainty. This will form the basis for the appreciation of simulation.

- Identifying how simulation works, together with its advantages over other approaches and its limitations.

- Explaining the importance of assessing and including the interdependencies between uncertain variables in a simulation model. This will lead to the construction of a model which allows the interdependencies to be considered and incorporated, through product-moment correlations.

- Illustrating how modelling dependencies can be achieved. Throughout the process, various statistical concepts will be discussed and it will be shown how they can be practically applied.

- Indicating how the RCM can be incorporated into a QRA model in practice. It demonstrates the value and effort of the complicated modelling process.

By the end of the research, a RCM is to be developed which will achieve the objectives defined above.

1.3 Contributions to knowledge

Upon completion of this research, the contributions to knowledge will be important in two academic areas, i.e. Statistics and Operational Research. They are as follows:

Statistical Contributions:

- Reinforcing the importance of modelling the interdependencies between uncertain components when simulation models are used.
- Discovering a distinctive way of formulating Pearson correlated sample values of a pair of continuous variables during sampling processes.
- Filling the gap between theoretical awareness of the significance of correlation and the actual practice of its use. In particular recognising the inappropriate use of rank correlations in many situations where the data are not ordinal and developing instead models of product-moment correlations.

Operational Research Contributions:

- Consolidating the advantages and limitations of choosing simulation as a means of carrying out risk analysis.
- Improving the reliability and precision of simulation output and exemplifying the sensitivity and confidence of using simulation methodologies.
- Showing the relevance and suitability of simulation and encouraging its wider use.

1.4 Thesis structure

A brief outline of the structure of this thesis is shown in Figure 1.1. The whole thesis is divided into four parts.

Part I Introduction and literature review (Chapters 1 and 2) provides an introduction to why this research is carried out, what are the purposes of doing this research, what are the aims and output of this research, and who it is to benefit. Chapter 2 emphasises a review of different approaches used when QRA is considered. It includes the advantages and disadvantages of each approach. It provides the theoretical background against which these approaches should be assessed, and how. In particular it defines and contrasts the two major measures of correlation (product-moment and rank correlation), and summarises key recent advances in the area of simulating correlations.

Part II Methodology (Chapter 3) provides the background to how this research will be carried out in order to achieve the aim and objectives of this research defined earlier in section 1.2. It will then identify the methodologies to be used in developing the RCM. In essence, then, this chapter provides the functional specification and the first part of the technical specification of the RCM. The second part is dealt with in Chapter 5.

Part III The Algorithms and the Computer Model (Chapters 4, 5 and 6) presented in the first part of Chapter 4 is the basis of the theoretical approach formulated for the RCM, together with a full illustrative example. The second part is the structure of the RCM as a computer model. Chapter 5 is the implementation of the theoretical approach into a computer based model where the theoretical approach is extended to include different types of probability distributions.

Chapter 6 is testing and verifying the validity of the Model and should ultimately demonstrate that the design and development of the RCM is appropriate.

Part IV Conclusion (Chapter 7) assesses the results generated from the RCM and discusses how closely the mathematical or statistical methodologies used in the model achieve the aim and objectives of this research. Further work extending from this research will be recommended.

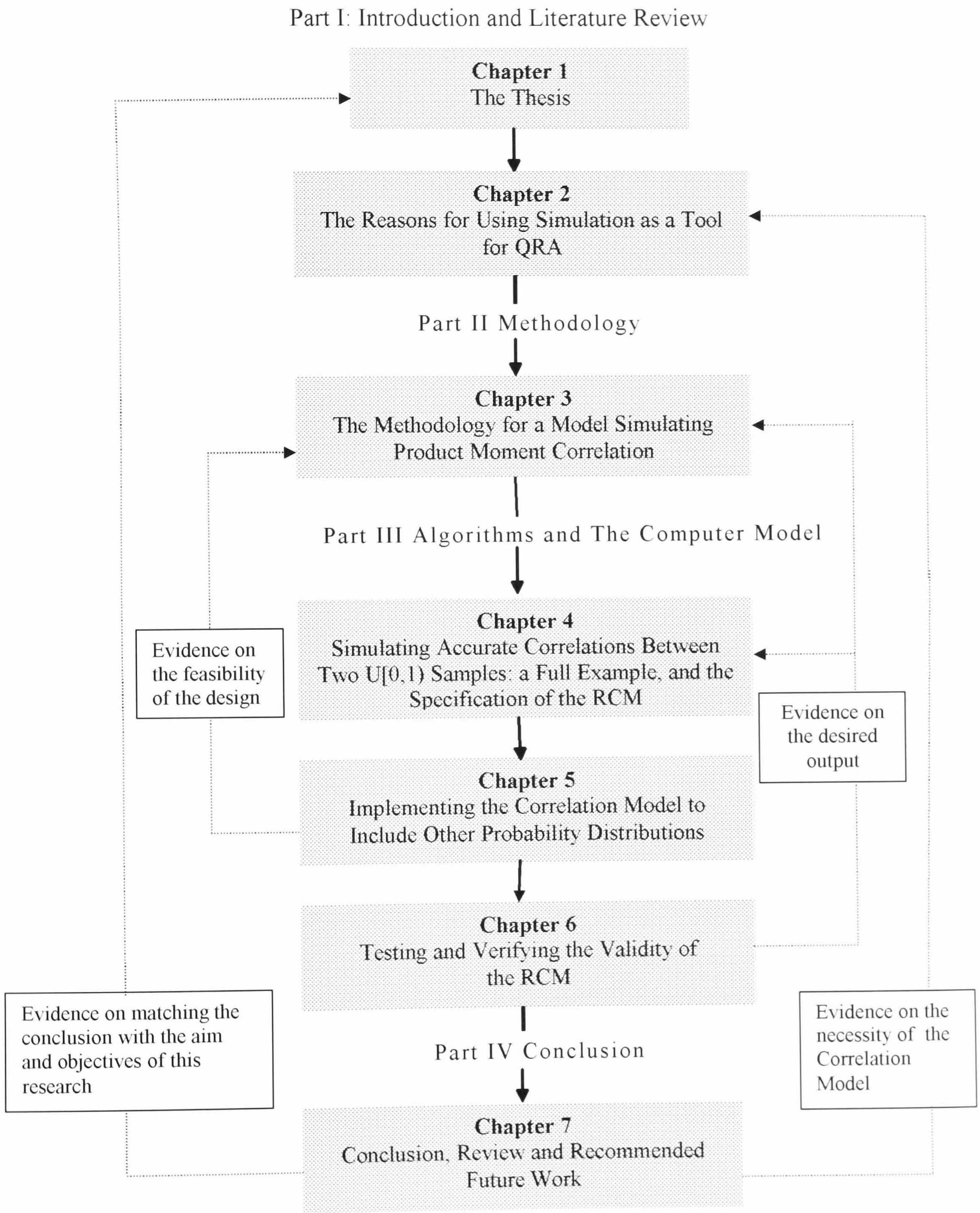


Figure1.1 Thesis structure

1.5 Overview of each chapter's contents

Within this section, the individual chapters of this thesis are overviewed. The aim, objectives and the outcomes of each chapter will be briefly identified.

Part 1: Introduction and Literature Review

Chapter 1: The Thesis

This chapter has provided an overview on this research in terms of the background to the research question, the research aim and objectives, and has also briefly described the evolution of the research. The research problem that has been identified in this chapter is that ignoring the interdependencies between uncertain components may result in unreliable output when carrying out risk analysis using simulation, and that there is a need to be able to model the relationship between two or more continuous variables using Pearson's product moment correlation. A methodology to solve this problem will be proposed. As a result, finding a way of incorporating interdependencies into a simulation model has become the aim of this research. By achieving the aim, this research will contribute to knowledge by filling in the gap between theoretical awareness of the problem and being able to model and incorporate appropriate measures of interdependency between continuous variables in practice.

Chapter 2: The Reasons for Using Simulation As a Tool for QRA

This chapter provides a chronological review of the existing literature. It starts with the definitions of risk, uncertainty and risk analysis, and proceeds to show how QRA is carried out in practice. This chapter categorises different QRA techniques, and each category will be discussed in detail in terms of its advantages and disadvantages and when it should and shouldn't be used. This leads to the rationale for this research and the reason why simulation is chosen to be the most appropriate risk analysis approach. It provides the incentive for investigating its limitations and overcoming them where appropriate.

The development and key aspects of discrete-event simulation are discussed in section 2.9.

The chapter concludes with formal definitions of the two major forms of correlation, examines some of their key properties, and reviews the relevant literature on the modeling in practice of joint distributions of several variables when their individual marginal distributions are known (or assumed), and their pair-wise correlations are predicated.

Part II Methodologies

Chapter 3: The Methodology for A Model Simulating Product Moment Correlation

This chapter presents the methodological considerations to be taken into account in this research. It illustrates the pre-requisite knowledge for formulating a model which is able to generate sample values of a pair of correlated continuous variables. This chapter presents step by step the mathematical and statistical formulations of how to generate a pair of correlated sample values from the assigned probability distributions of the two variables, and which have a measure of interdependency via a defined product moment correlation coefficient. It also provides a strong justification for using Latin hypercube sampling in simulation, rather than the various traditional Monte Carlo approaches.

Part III The Algorithms and The Computer Model

Chapter 4: Simulating Accurate Correlations Between Two U[0,1) Samples: a Full Example, and the Specification of the RCM

An illustrative example will be given to clarify the explanation of the algorithm formulation discussed in Chapter 3. As the RCM will be transformed into a computer based model, this is where the second part of this chapter also concentrates on the structure used to achieve the aim of this research from the perspectives of formulation, design and development.

Up to this point, the research is limited to variables which have standard Uniform distributions (i.e. U[0,1)), but this will be extended in chapter 5. These formulations are implemented as a theoretical model, and the testing of the output from this model will be compared later in Chapter 6 with that from a commercial simulation package currently available.

Chapter 5: Implementing the RCM to Include Other Probability Distributions

Within this chapter, the theoretical approach designed in the previous chapter with the extended scope to include other probability distributions will be implemented as a computer model. The structure of the completed RCM will be presented as a summary flow chart with fuller details of each constituent part.

Chapter 6: Testing and Verifying the Validity of the RCM

This chapter is concerned with documenting the testing process. The descriptive statistics calculated from the RCM will be tested using Microsoft Excel. This is to ensure that the programming for producing the descriptive statistics is free from human error. Descriptive statistics form a list of measurements on the sample values. They are presented in the model so that they can be used to check and compare the sample values generated with the specification of what was required.

Once the model has been specified and then developed, it has to be verified that it is functioning as desired. In this chapter, the descriptive statistics calculated from the sample values generated from the RCM are investigated to ensure their acceptability. This chapter will support with evidence the claim that the formulation of the model has been appropriately designed, and will demonstrate an overwhelming improvement in the modelling capability of product-moment correlations compared with standard commercial packages.

Part IV Conclusion

Chapter 7: Conclusion, Review and Recommended Future Work.

This chapter sets out the overall conclusion on how closely the research output meets the initial aim and objectives. A summary of the problems faced during the research and the limitations on the research output is examined. Recommendations for further work extended from this research are suggested. For example, extending the analysis to cope with correlations between more than two continuous variables, and developing processes to model more accurately other measures of a variable such as its skewness and kurtosis.

Summary

Within this first chapter, the problem of ignoring the correlation between variables in a simulation model has been recognised, and has led to the definition of the research question. This is followed by the research aim, to specify, formulate and develop a Pearson correlation model that can be incorporated into a simulation model. During the process of modelling, several desirable objectives have arisen, and these have been listed in the chapter. When they are all achieved, the contributions to the academic area of statistics and operational research will be significant.

The structure of this thesis is divided into four main parts, each containing one or more chapters, and it is presented in a flow chart in Figure 1.1, the contents of which have been discussed in overview detail.

Chapter 2: The Reasons for Using Simulation As a Tool for QRA

2.1 Introduction

The objectives of this literature review chapter are two-fold. The theme of the first part begins by exploring the debate about risk and uncertainty, especially their meaning in the eye of Operational Research.

Such terminology is not only the prerequisite for appreciating the need of risk analysis during, for example, a capital budgetting process, but also provides the comprehension for reviewing the emergence of using simulation as a tool for quantitative risk analysis (QRA).

While the theme of the second part of this chapter is to review the emergence of using simulation in the QRA process, the different methodologies used in practice when uncertainty is taken into account are examined, with capital budgeting again being the vehicle for explaining or interpreting theses methodologies in many cases. Hull (1980) and Smith (1994) were of the opinion that capital budgetting decisions are among the most important of all management decisions.

After evaluating the advantages and disadvantages of using each methodology, the reasons for using simulation as a method for carrying out QRA becomes clear. These methodologies are summarised from different techniques used when uncertainty is considered.

Similarly, at the end of this chapter, the reasons for nominating the particular forms of the research question and the aim and objectives of this research in Chapter 1 are justified.

2.2 The environment of capital budgetting

Capital budgetting is one of the most risky elements in the finance function due to the uncertainty in prevailing economic conditions (Van Horne, 1995). It is highlighted as a vehicle here because the recognition of risk as an important component in capital budgetting decision-making has long been identified (Brookfield, 1995). The practice of using simulation in capital budgetting applications has been growing in recent years.

Chansa and Mount-Campbell (1991) suggested that further research in the field of capital budgetting is required and it should be concentrated on getting high quality project cash flow information based on uncertainty economic conditions. This can be done by developing appropriate tools for handling and reducing the riskiness of the investment decisions.

The techniques used in including considerations of uncertainty are known as risk analysis techniques (Smith, 1994). This research is specifically focused on the application of risk analysis via simulation in situations where the variables are continuous. Consequently our main attention will be directed towards quantitative risk analysis (QRA) from Chapter 3 on.

Any investment appraisal technique that fails to take consideration of the cash flow information is most likely to lead to incorrect conclusions and erroneous recommendations (Brookfield, 1995).

Vose (1996) asserted that an understanding of the techniques used to carry out risk analysis has not been matched by a corresponding growth in its popularity amongst businesses and government agencies, although the relatively recent emergence of risk analysis tools such as @RISK and Crystal Ball have enabled this delayed growth of use to begin.

The above has certainly encouraged the needs for this research to look into the most popular approaches used in developing quantitative risk analysis models and help fill the gap of knowledge and also credibility in QRA. In the area of capital budgetting, for example, the result of this research might be to enable the development of a more comprehensive and useful project cash flow information tool that will improve the quality and confidence in the output which is used to aid decision making.

2.3 The nature of risk

Uncertainty and risk are the main components of any activities. They are not only limited to our private lives, they also occur in virtually all business decisions.

Most of us have learned to live comfortably with day to day uncertainties and to make choices and decisions in their presence. When there is no great impact from a failure then commonly the uncertainty is simply ignored (McCray, 1975, and Morgan and Henrion, 1990). Nevertheless, Vose (1996) suggested that uncertainty and risk need to be understood so that rational decisions can be made.

2.3.1. The meaning of uncertainty

As early as in Rowe (1977), through Ritchie and Marshall (1993), and up to Vose (2000), the definition of uncertainty has remained consistent. It can be concluded that uncertainty arises from one's imperfect knowledge about the past and/or doubt about the future, specifically the proposed decision and its possible consequences. This is illustrated in Figure 2.1:

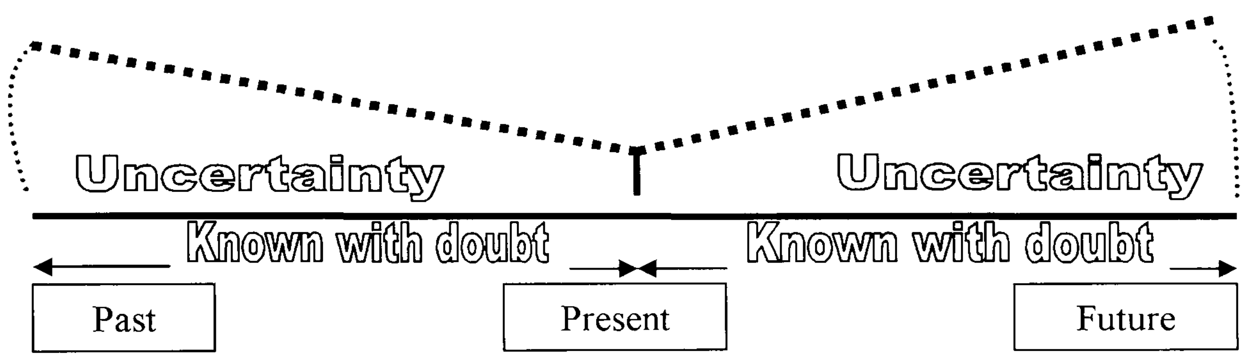


Figure 2.1 Uncertainty in time scale

Each person's imperfect knowledge could arise from not knowing precisely what had happened in the past and being unsure of what will happen in the future. The further the time frame is from the present, the more difficult it is to sketch the picture of the past or to predict the future with any confidence.

It is more difficult to control the outcome of the event when one lacks information surrounding the future event with certainty, and this lack of certainty is, unfortunately, virtually always present. Rosenhead (1989) stated that it is dangerous to "attempt to pre-take the future". He focussed on the common practice of assigning probabilities of occurrence to the individual outcomes of events, whether they be measured on discrete or continuous scales. Wherever possible such estimates should not be used.

However, although clearly it is sensible to pay careful attention to the justification of choices of probabilities, including probability distribution functions (p.d.f.s), there is often no alternative because QRA has to be carried out. Hence the relevant forecasts, etc., which arise from this process should never be regarded as deterministic as such information can only be predicted with at least a level of doubt.

Example

It help to illustrate the definition of uncertainty in a simple example. Suppose an entrepreneur is thinking of launching a new product into the market. He accepts that there will be two possible outcomes: either this new product will be accepted or rejected by the market.

Action	Outcomes
Launch a new product	Accept Reject

Here there is an uncertainty inherited in the action. This is because there are at least two possible outcomes, and the entrepreneur's uncertainty comes from his lack of information about how successful his launching would be. Actions open to him would include trying to estimate the probabilities of acceptance or rejection, perhaps by reviewing similar launches in the past, and evaluating their outcomes.

It is also interesting to observe that a person, particularly a layman, will try to express his uncertainty as to the outcome of future events through the use of words such as 'probable', 'possible', 'expected', and 'likely'. Unfortunately some of the general words used in this context have very specific meanings in parallel contexts. E.g. 'expected'.

Uncertainty in a project

Uncertainty arises when any of the events which constitute a project have wide ranges of possible outcomes as a result of imperfect knowledge of the events. In some cases these potential outcomes are mutually exclusive; in others there may well be a degree of interdependence.

Such events could be a change of the taxation rates, changing demand for services/products, or a variation in costs and/or revenues, for example, and taken together they are highly influential on the project profitability.

When any one or all of the events have many different possible outcomes, this will make the overall outcome of the project highly uncertain.

2.3.2. The meaning of risk and its relationship with uncertainty

Risk can mean different things to different people (Cochrane, 1992). Singhvi (1980) said that "risk, like beauty, lies in the eyes of the beholder".

Given that the word 'risk' is used in many different contexts with an equally wide variety of definitions, if we do not have a particular context in mind when asking people about risk, they will make up their own contexts based on their own experiences, beliefs, habits, etc.

These contexts will be as immensely varied as each person's own experiences and concerns. People tend to relate risk to specific situations where there are particular stimuli considered being dangerous.

2.3.3. Risk as a result of uncertainty

Hertz and Thomas (1984), Cooper and Chapman (1987), and Eschenbach (1996) gave the opinion that risk is something *concerned* with uncertainty and also *resulting* from uncertainty.

Event + Uncertainty = Risk

The above expression implies that when there is uncertainty, there is risk. When an event (or activity) involves uncertainty, a risk arises from the decision.

From the earlier example in section 2.3.1. the uncertainty is whether the new product will be accepted by the market if it is launched. If the action is undertaken, it is saying that the entrepreneur is taking the risk of making the decision or the decision is risky.

2.3.4. Risk as the impact upon a decision maker

Ho and Pike (1991) raised the opinion that risk is a measure of the consequences that impact on projects from the occurrence of an event. As a result risk is then a measure of how a particular project will impact upon the investor (or decision maker).

The philosophy above on risk has two essential prerequisites: uncertainty and loss. Risk is then used to denote that the decision maker is uncertain as to the precise outcomes of the investment decisions, which involve the possibility of undesirable consequence or loss.

Less formally, Cavinato (1990) said that risk is the chance that a project you are sticking your neck out for will not live up to an estimated outcome.

So, if the outcome of an action is uncertain or uncontrollable and may cause some loss (e.g. of money, human life), the action is risky (Indoinc.com, 1999).

2.3.5. Risk as the notion of probability

Risk as the notion of probability is not new to the academics. Singhvi (1980), for example, recorded that in 1975, a survey was carried out by Petty, Scott and Bird on 109 industrial corporations. A question was asked to management, '*What is meant when you say an investment proposal is risky*?' The results of this survey are summarised in Table 2.1 following.

Definition of risk	% of total responses
Probability of not achieving a target return	40
Variation in returns	30
Payback period uncertain	10
Uncertain market potential	7
Entering an inexperienced area	5
Success ratio (potential gain/potential loss)	4
Miscellaneous	4

Table 2.1 Management's definition of risk (Source: Singhvi, 1980)

The survey showed that the definition of risk in the eye of management as the probability of not achieving a target return, for example making a loss, was identified by 40 per cent of the responding executives.

Next, 30 per cent of the respondents were principally concerned about variation in returns (although nothing is suggested in the analysis about the direction or directions of such variation). Only 10 per cent of the respondents define risk in terms of the payback period, which is the length of time before the cumulative expected return is at least equal to the costs incurred to date.

Although this survey was carried out a long time ago, the conclusion produced from the survey, i.e. risk is the probability of not achieving the target, remained unchanged according to Ritchie and Marshall (1993).

2.3.6. Stems of risk, illustrated via capital budgetting

Uncertainty arises from imperfect knowledge about future event and risk is the consequence of uncertainty. It is important to identify – and hence try to control where possible – the factors which are contributors to this uncertainty.

For example, in capital budgetting the major uncertainty comes from the predicted data, so that any decision based on decision criteria is recognised to be risky.

Cavinato (1990) and Ho and Pike (1991) explain that the uncertainty inherent in the predicted data stems from three areas, as in Figure 2.2 below.

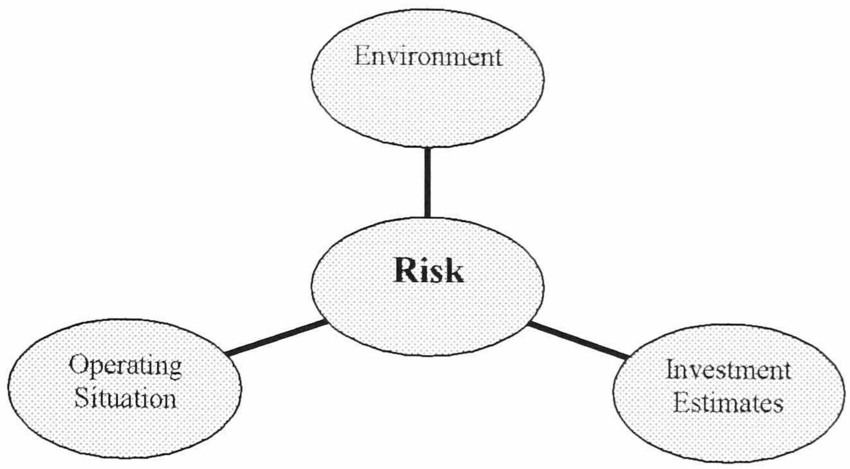


Figure 2.2 Stems of risk in capital budgetting

The business environment might change for the worse. For example the underlying economic environment might deteriorate through considerable instabilities in inflation levels and exchange rates following currency flotation.

Similarly, the operating situation could deteriorate, for example if the estimated schedule is not achieved on time. Also the investment estimates might turn out to be wrong. For example, the estimated revenue and expenditure might be overestimated and underestimated respectively.

In conclusion, when decision makers frequently have to confront the possibility of making the wrong decision and experiencing 'negative' outcomes such as financial loss, Cooper and Chapman (1987) claimed that risk analysis has the greatest immediate acceptance in the area of investment project appraisal. It is therefore important to realise that the greater the understanding of the precise nature and level of the risk the better the decision will be and, ultimately, the firm's performance (Hertz and Thomas, 1984; Cooper and Chapman, 1987, and Ho and Pike, 1991).

2.4 Risk analysis

Risk analysis was jelled into a distinct discipline around 30 years ago (Tversky, 1974). It is increasingly demanded as part of capital project justification (Cavinato, 1990). As stated earlier, Hull (1980) and Smith (1994) claimed that capital budgetting decisions are among the most important of management decisions.

Hillier (1962) advised that the amount of risk involved is often one of the important considerations in the evaluation of proposed investments. Decisions must be made in the face of uncertainty (Eschenbach, 1992; Randhawa and Douglas, 1993). Very often they involve a relatively large commitment of a company's resources and are instrumental in shaping its whole future. To a large extent the expenditures that project decisions involve are irreversible. Mistakes in capital investments not only affect immediate cash flows but also the operation of the business and future cash flows for years to come. E.g. The huge cost-over-run of the building of the new Scottish Parliament.

Thus, the need to manage uncertainty is especially desired for large projects that have not been executed before and therefore involve much uncertainty in project schedule and project cost (Lucey, 1968; Mott and Tumay, 1992). Bierman (1986) confirmed, via a survey of US senior financial officers, the challenge of handling risk was one of the most prominent problems in capital budgetting practice. This has encouraged a research in finding or developing a tool for evaluating risk in the capital budgetting process.

2.4.1. Risk analysis definitions and its usefulness

Hertz and Thomas (1984) declared that risk analysis is used when developing a comprehensive understanding of the risk associated with a variable is necessary.

Within this research, this declaration will be extended to cope with several variables, together with their inter-dependencies. "Risk analysis is a process of identifying and evaluating risk factors and a study of the likelihood that an event will produce an unwelcome outcome or adverse effect of what may be and the impact should a failure occur" (Swaney, 1996; Quality Assurance Review Guide, 1999). Bodily (1992) observed that the risk analysis process is primarily through estimating probabilities which involves collection and analysis of data, originally to estimate human morbidity or mortality, now expanded to range across many areas, from ecological health to economic well-being, etc.

Risk analysis therefore estimates a range of possible results of a proposed investment decision based on the given input data and seeks to quantify the probability that the overall result will be in a specified range. It simulates the effects of the uncertainty surrounding key variables entering into the evaluation on the returns one is likely to achieve (Singhvi, 1980).

Consequently risk analysis is used to examine the possible future outcomes before approving an investment proposal, a new product, or a future corporate strategy (Karady, 1985). For example, in evaluating a capital project, a company carries out a risk analysis to determine its financial risk in making the investment. The approach might incorporate a cash flow model, and the risk analysis might involve a simulation of the uncertainty in the net present value cash flow or other financial performance measures.

Risk analysis therefore efficiently provides risk managers with quantitative evidence which will enable them to:

1. break down the complex problem into smaller and more manageable sub-problems, each of which can be analysed;

2. measure exposure to extreme events. For example drastic market moves, or large changes in interest or exchange rates;

3. reduce the risk of the project, if it is not acceptable, either by diversifying, risk sharing, or contingency planning to protect against unwanted scenarios (Balson et al, 1992); and

4. provide better ways for individuals and groups to reduce hazards or cope with their efforts (Phoa, 1999).

2.4.2. The significance of risk analysis in the capital budgetting decision

The researches of Richards and Contesse (1975); Hosseini (1986); Cozzolino (1979); Coats and Chesser (1982); Hertz and Thomas (1984); Karady (1985); Ho and Pike (1991); and Chadwell et al (1996), all concluded that firms recognised that risk analysis was critical in the proper evaluation of capital projects. Management relies heavily on risk analysis techniques for evaluating complex strategic projects, so that corporate success can be partly attributed to the use of such approaches.

However, varying degrees of risk among projects should be taken into account (Eschenbach, 1992), so that the greater the risk in the outcome, the greater is the case for using the formal techniques of risk analysis. Eynon (1988) claimed that one of the seven deadly sins in the decision making process is to ignore risk analysis completely.

Hence, it is highly desirable for the success of the company during project planning to have as accurate financial estimates as possible. The system of generating accurate estimates must be established and maintained. In the many changes that take place within companies, the importance of generating accurate estimates cannot be overlooked.

The contribution which risk analysis can make is to help managers' thinking processes, and this is done in the first instance by forcing them to confront the structure of the decision problem in a relatively unemotional manner.

After the problem has been defined, Karady (1985) specified that evaluation of the cross-impacts or joint impact amongst the uncertain variables is vital in the risk analysis. By doing so, viable options can be identified and decisions makers eventually understand the risk associated in the project, and appreciate why one course of action might be more desirable than another.

His thought eventually has become the stem of this research. This aspect of 'cross-impacts / joint impact' referenced by Karady is thus reflected in the examination of product-moment correlations in the RCM.

2.4.3. The application of risk analysis

> "who is at what kind of risk, when, where, with what effects, from what causes, with whom, responsible for, by what instruments, in what value context, and at what costs and benefits for its management?"

Coates (1994) identified the general risk analysis question above as the starting framework for anyone quickly to develop a comprehensive overview of any situation with regard to any risk they are concerned with.

The principle of risk analysis may be used in various application areas (Quality Assurance Review Guide, 1999). For example, in the United States, and to a more limited extent in European countries, risk analysis emerged during the 1980s as a major methodology for regulatory policy-making (Brod, 1992). Public agencies found themselves increasingly influenced by the impacts of national and state legislation, budget constraints on operations, new regulations and growing demands for resources.

As this changing environment became more complex, it required the adoption of systematic approaches for evaluating the consequences of alternative management policies and external events.

Risk analysis can be equally applied to both qualitative and quantitative evaluations of the risk arising from some activity (Bodily, 1992). It is tabulated in the user guides of two popular commercial simulation and risk analysis packages, Crystal Ball and @RISK, that risk analysis can be used in different application areas.

Both listed areas such as those below, for each of which an illustrative example has been given.

Oil and gas - Texaco uses risk analysis to forecast inventory requirements and optimise production levels.

Project management - risk analysis allows Hewlett-Packard to bring printers to market on-time.

Finance – many companies, for example, the ProVise management group, use risk analysis to optimise portfolio profit.

Negotiation Litigation – Pacific Bell called upon risk analysis to help negotiate financial settlement in 1994.

Business planning – risk analysis boosted the new venture planning for recreation markets of a major American company, ExpertCorp.

Costs management – 3M's use of risk analysis improves unit cost estimates.

Forecasting – Risk analysis is used for forecasting prison populations by consultancies such as Fentress Inc.

Environmental – Alcean uses risk analysis to determine the environmental damages that are caused by people who consume drugs in the United States and throughout the world.

2.5 Accounting for risk in decision making

The importance of considering uncertainty, in capital budgetting or elsewhere, has widely been recognised by both practising managers and the academic community (Hull, 1980). Various techniques are used by decision-makers to cope with the risk associated with a proposed investment or project.

These methods can be categorised into qualitative or quantitative techniques (Singhvi, 1980; and Smith, 1994). Qualitative techniques will only be mentioned in brief when it is necessary as part of the main discussion in this thesis which is essentially that of QRA (i.e. Quantitative Risk Analysis).

Qualitative techniques are used to distinguish the possibility of a risk occurring in a linguistic manner (Baker et al, 1998). For example, a risk might described as low if that risk is unlikely to occur. Qualitative techniques are usually employed at the beginning to identify and rank risks. Those risks with a high or intermediate rank may then be further analysed through quantitative techniques.

Examples of qualitative techniques to handle risk arising from uncertainty are various forecasting and problem structuring techniques, such as scenario writing, cross-impact matrices, robustness analysis, cognitive mapping, relevance trees, professional judgement, personal experience or brainstorming (Hanke et al, 2001).

These techniques are prone to inconsistencies because they are dependent on the experience of the analyst allied to the judgements, and thus are subjective.

On the other hand, quantitative methods of coping with risk are normally computationally based and utilise relative frequencies during the estimating of numerical probabilities of the consequences and likelihood of identified risks (Singhvi, 1980, and Baker et al, 1998). Example techniques here are decision trees, portfolio theory, simulation, and risk-adjusted decision making methods. The results of a quantitative technique are compared against company criteria and decisions are made as to whether the risks are acceptable or not.

Data used in quantitative techniques are either obtained from historical databases or are estimates, and so they contain some element of uncertainty, due to the possible use of subjectively attained values. The level of judgement required for each method used in quantitative techniques is discussed in the next section.

2.6 Approaches: quantifying uncertainty

Quantifying uncertainty is the first distinctive feature of risk analysis (Singhvi, 1980). Various methods used to cope with uncertainty all essentially break down into three recognisable approaches, i.e. point analysis, scenario analysis and simulation analysis as shown in Figure 2.3 below, together with some examples.

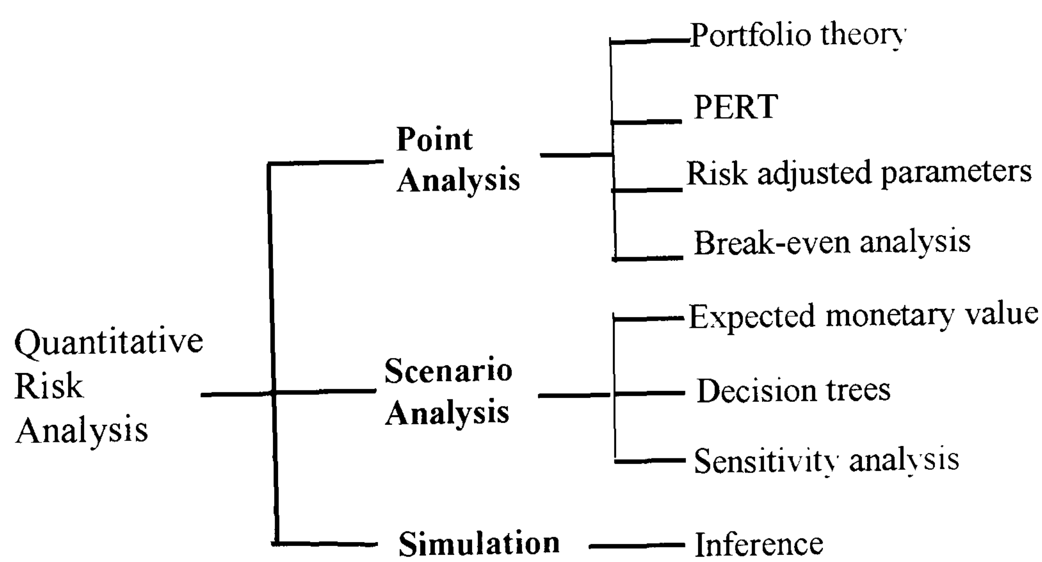


Figure 2.3 Examples of approaches used in coping with uncertainty

Authors, including Hertz (1964), Singhvi (1980), Gapenski (1990), Smith (1994), and Nanda and Miller (1996), claimed that, traditionally, the risk analyst tried to capture this uncertainty either using point analysis or scenario analysis, or sometimes both in the cash flow calculation such as NPV cash flow calculations or the IRR for project investment appraisal. However, when there are limitations on using the traditional methods to account for uncertainty, simulation is a better method in the project appraisal process, as explained below.

2.7 Point analysis

The traditional approach in developing the cost and schedule components of a project has been to create single point estimates and schedules with single point completion dates (Wendling and Lorance, 2000).

Single point modelling involves using a single 'best guess' estimate, i.e. the value which one thinks is most likely to be achieved, of each variable within a model to determine the model's outcome(s), including the uncertain variables.

Gapenski (1990) described how, in a typical capital investment feasibility study, the analyst makes point estimates of the relevant component cash flows and then uses these values to forecast the expected profitability of the project. In this sense the model is deterministic. This all fits in nicely into the traditional decision process shown in Figure 2.4.

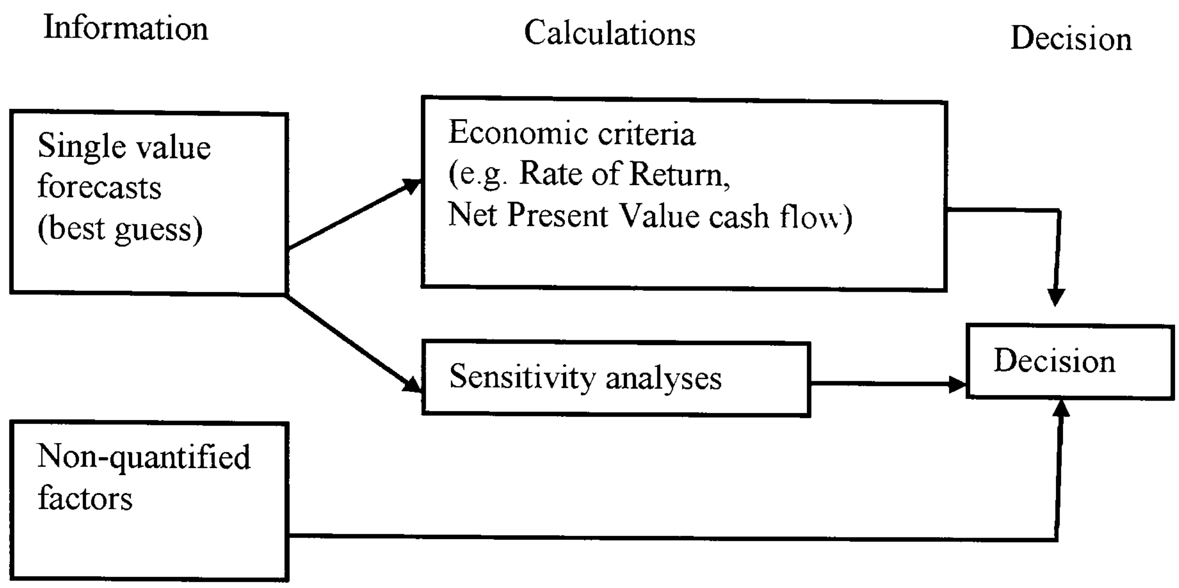


Figure 2.4 The traditional decision process.

(Source: Cooper and Chapman (1987, p208)).

For example, suppose the analyst projects the best estimate on annual inflation at 3.5 percent for the next 5 years. For the same period, the analyst estimates that the capital costs will increase by seven percent a year and the other (operational) costs will increase by eight percent annually.

Even if these projections are reasonable by historical standards and/or professional judgement, they are still merely estimates. The analysis often stops here, with high potential for incorrect interpretation, for example claiming that the profitability of the new investment is known with certainty and there are no risks involved. To correct this problem, the analysis must be extended to incorporate the uncertainty inherent in the project itself.

2.7.1. Risk adjusted parameters

When there is a limitation in their ability to predict future estimates with any degree of precision, some analysts try to manipulate the "best guess" estimate by adjusting the values for each input variable or uncertain value. This will follow the 'Rule of Conservatism'. I.e. When estimating revenue, estimate low; when estimating cost, estimate high. This is often the first line of defence a project proposal goes through when it gets passed along to upper management (Cavinto, 1990). There are various parameters which can be adjusted to cope with the risk or uncertainty associated with a proposed investment.

2.7.2. Example of risk adjustment: cash flows

Decision-makers often reduce the best estimates of future cash inflow. For example, an analyst may estimate the capital cost of a project at £500,000. But because of the uncertainty associated with this estimate, and the risk that the cost may turn out to be higher, they may use for evaluation purposes and to avoid underestimating the future cost, a value which is written up by, say, 10 percent, at £550,000. However, this clearly could lead to the project being deemed to just be uneconomic, whereas using the 'correct' estimate, £500,000, might lead to a positive economic assessment, so an additional source of risk will have been introduced inadvertently by this effort to avoid risk. This approach is frequently used by 'Risk Adverse' companies.

Clearly such additional risk factors must be stated openly, lest they should not be recognised to be as such at some time in the future.

2.7.3. Example of risk adjustment: discount rates

The risk adjusted approach may even be applied to the discount factor which is the rate used to calculate the present value of future cash flows in the discounted cash flow analysis, e.g. NPV cash flow calculation. The calculation of a NPV cash flow requires the selection of a discount rate, so that the decision may be to raise the discount rate for risky projects in relation to the overall cost of capital. For example, the raising of a risk-adjusted discount rate from, say, 13 percent to 18 percent, might be used to account for risks. Again, unless the reasons for using that value are explained, the decision arrived at from the evaluation might be very different.

Certainly, the use of unrealistically high risk-adjusted discount rates would tend to reduce the risk of a company making an unproductive investment. Indeed, if there were an unlimited number of highly profitable investments in the industry waiting to be developed the use of this practice might be justified. If a project is still acceptable when discounted at this larger rate, this is fine.

However, within an industry like the petroleum industry which is highly competitive, the use of risk-adjusted discount rates might lead to the rejection of acceptable investment opportunities.

The conclusion is that the implementation of different discount rates implies that managers are incorporating risk in their long-term investment decisions.

2.7.4. Example of risk adjustment: the payback

The risk adjusted approach can also be applied when the analyst does not use discounted cash flow analysis of investment opportunities. Their basic decision criterion is then the Payback – the time required to return the original investment in undiscounted terms. The basic concern is: "*How soon will we get our money back?*" In this case, the objective of accounting for risk is frequently a move simply to reduce the acceptable payback period for screening purposes, for example from 5 years to 3 years. In other words, the decision-maker may define a lower payback period for risky projects than the target payback period for normal risk projects.

From the above analysis, adjusting the input values in the cash flow calculation to account for risk may appear at first sight to be very convenient, but it can return very misleading results and biased decisions. Gapenski (1990) pointed out that such cases the projects are being analysed subjectively and conservatively.

Van Rensburg (1990) showed how adjusting the input parameters on the discounted cash flow decision criterion could be significant, and he therefore concluded that even minor adjustments to variables could have a significant effect on these criteria. Adjusting the input parameters not only cannot solve the problem of uncertainty, but a certain level of bias will be indirectly incorporated into the process by so doing. Certainly the analyst will have lost a degree of control over the situation.

2.8 Sensitivity analysis

Using the risk adjusted parameters method on the point analysis by changing the best-guess estimate of the variable's subjectively builds risk into the evaluation. Brod (1992) maintained that the limitation of a forecast having a single expected outcome is clear: while it may provide the single best guess, it offers no information about the range of probable outcomes.

Since conventional analysis had failed to give a satisfactory result using single best estimates in the project appraisal process, Pouliquen (1970) claimed that the most natural way to deal with this situation was to make a sensitivity analysis.

A sensitivity analysis is also frequently referred as a 'scenario analysis', in which case it should not be confused with 'scenario writing', which is a qualitative method mentioned in section 2.5 above.

Sensitivity analysis is used to see what would happen if other values of the input data were substituted and to examine the effects on the profitability criterion function of changes in the values of the key variables. One form of sensitivity analysis is to see how far revenues would have to drop or savings diminish until the minimum rate of return is reached (Cavinto, 1990). However, McCarthy (1994) argues that scenario analysis is merely an extension of the point analysis.

By combining several point estimates, the analyst hopes to "bracket" the uncertainty in the projection. A particular case of sensitivity analysis is to assign minimum, best guess and maximum values to the key input variables and compute the corresponding values of the decision criterion, thus providing a range of possible results. These various combinations are commonly known as 'what-if' scenarios (Vose, 1996).

For example, Table 2.2 concerns the profitability of a proposed petroleum project, and has been broken down into five separate factors. Three values of each factor's value are included, i.e. minimum, best guess and maximum. In this table the abbreviation "b" represents "barrels". Since there are five factors and three values per factor, $3^5 = 243$ possible 'what if' combinations could be produced. This type of analysis improves the point analysis. However, there are a number of criticisms from different point of views, which are discussed below.

Key Factors	Minimum	Best Guess	Maximum
Capital Expenditure (£mm)	305.0	332.0	378.0
Exchange Rate ($ to £)	1.75	1.82	1.95
Oil Price ($ per b)	24.0	38.0	48.0
Recoverable Oil Reserves (mmb)	44.0	52.0	80.0
Production Rate per Day (mb/d)	29.6	37.2	43.6

Table 2.2 Example of petroleum project factors

When sensitivity analysis was first introduced to replace point analysis, a number of researchers supported its use. For example, Rappaport (1967) and Hertz (1979) emphasised that sensitivity analysis is a logical adjunct to deterministic capital budgetting, particularly as a means of developing a better initial understanding of the nature and impact of risk. Hull (1980) and Chapman and Ward (1997) claimed that all effective quantitative modelling requires sensitivity analysis, so the analysts and the user of analysis can understand the relative importance of the components the analysis uses. Sensitivity analysis enables the most important parameters to be identified for further analysis, more detailed monitoring, or more sophisticated forecasting.

Sensitivity analysis using estimated single values is probably the most common method used in the quantitative economic evaluation of a venture. The method determines the relative sensitivity of a particular parameter's value and indicates those parameters which have the most influence on the measuring criterion.

However, in the above example shown in Table 2.2 it is likely that there is too large a set of scenarios to have any practical use. As described in Singhvi (1980) and Mackenzie (1989), sensitivity analysis provides management with answers to a wide array of 'what if' questions. Beyond some point, however, this rather mechanical exercise becomes less useful.

This particular case of sensitivity analysis suffers from two important drawbacks, and other more general cases suffer from the second of these drawbacks. Firstly, the use of only three values of each factor. But what are the three values? Since there is no standard way of choosing the three values for the scenario analysis, they will vary from person to person and, therefore, bias in choosing those values will most probably occur. As concluded in Coleman et al (1995), the creation of multiple scenarios by arbitrarily varying key assumptions to account for future investment uncertainty is not well-suited for estimating risk in an economically meaningful fashion.

Secondly, no recognition is given to the fact that a value close to the best guess value is much more likely to occur than the minimum and maximum perceived values, and so the three values are given the same weight. Therefore, while scenario analysis improves the point analysis by giving the best and worst case and gives the analyst a wider margin of error, it is still does not solve the problem in the point analysis. i.e. The likelihood of occurrence of a particular outcome of an event is not provided.

Bennett et al (1970), Lucey (1968), and Wagle (1978), show how sensitivity analysis lets the decision-maker concentrate efforts on refining those estimated values of parameters having the greatest effect on the rate of return. However, the result is still a single statistic providing neither indication of risk nor the degree of interaction between parameter values. It does not indicate the likelihood of obtaining this particular outcome.

Ironically, the more combinations of variables one tries, the less clear the picture of the project may well become.

Thus sensitivity analysis does not in itself assess the risk of an investment alternative, although it usually identifies potential sources of risk. To measure risk, we have to incorporate probability estimates of the 'what-ifs' occurring.

McCarthy (1994) emphasised that decision analysis should really be able to give the whole picture of how likely are the occurrences of the possible outcomes of a project in terms of probability. For example, what is the probability of earning £50,000 from a specific investment?

By highlighting a few key variables from the many project variables, sensitivity analysis helps focus the limited time and effort available for evaluation in the most productive way. Perhaps most importantly, sensitivity analysis plays a useful role in the evaluation process by providing a bridge between single-point appraisals of expected value and probabilistic risk analysis. This principle weakness of scenario analysis or sensitivity analysis, i.e. the failure to provide a probability of occurrence of the outcomes, is overcome by using simulation with probabilistic analysis, and this is discussed in the next section.

2.9 Simulation analysis

The Term Simulation

Cooper and Chapman (1987) and Nanda and Miller (1996) termed simulation as a technique that imitates the operations of a real world system as it evolves over time.

The explanation behind this is that risk analysis models using simulation manipulate probabilities and probability distributions in order to assess the combined impact of risks on the project. As pointed out in section 2.6 above, the first distinctive feature of risk analysis is quantifying uncertainty. Here, the second distinctive feature of risk analysis is simulating the outcomes. It simulates the effects of the uncertainty surrounding key variables entering into the evaluation on the returns one is likely to achieve.

This is important, as described by both Pouliquen (1970) and Singhvi (1980), who wrote that risk analysis should be used to estimate a range of possible results of a proposed investment decision based on the given input data and to state the probability that the overall result will be within a specified range. In other words, a forecast is obtained for a variable of interest in the form of a probability distribution.

Risk analysis, by using simulation, is similar to scenario analysis in that it generates a number of possible scenarios. However, it goes one crucial step further by accounting for every possible value that each variable could take and weights each possible scenario by the probability of its occurrence (Vose, 1996).

Mott and Tumay (1992) suggested that whether the decision is to avoid risks or to take calculated risks, one thing the decision maker must do is "to buy insurance with an investment in simulation".

In this case, risk analysis should not only provide a tool by which risks which can impact on project estimates of costs, schedule, and production can be quantified, the joint impact of these risks can also be examined. The most important is the identification of the definite perception of what overall risk really exists.

2.9.1. The development of simulation

Silbergh (1972) noted that before the advent of simulation, the decision-maker dealt with uncertainty in qualitative ways by making conservative forecasts or by using a risk-adjusted discount rate, or both. However, as shown above, one conclusion to be drawn is that risk analysis should eliminate the need for restricting one's judgement to a single 'best', 'worst' and 'most likely' evaluation.

An early application of simulation to the analysis of the project investment appraisal was described by Hertz (1964). Before that, manual simulations, such as moving troops through the field, or playing a board game, were used. However, this can be very time consuming and with one or two simulated outcomes provides very little information on which to base a decision. With the increasing availability of faster and more poweful computers, and better understanding of quantitative modelling, simulation has become a very popular approach in recent years for the analysis of business problems.

Evans and Olson (1998) confirmed that simulation today is widely accepted in the world of business to predict, to explain, to train, and to help identify optimal solutions. Simulation is used extensively in manufacturing to model production and assembly operations, develop realistic production schedules, study inventory policies, analyse reliability, quality, and equipment replacement problems, and design material handling and logistics systems. It is used in designing and evaluating computer and communication networks and scheduling resources in complex projects.

Simulation also finds extensive application in both profit-seeking service firms such as financial and retail companies, and in non-profit service organisations such as health care, government, and education. These applications involve the study of customer waiting-line behaviour, evaluating surgical schedules, and designing efficient work flows in offices, for example.

For instance, simulation models can be used by a bank to help identify the number of tellers required to maintain a specific level of customer service as measured by waiting time or line length. Coleman et al (1995) observed that statistical market risk assessment uses probabilistic cash flow analysis as an objective means of measuring and analysing market-driven credit risk for a commercial real estate property, security, or pool of securities.

2.9.2. The simulation process

Simulation calculations are made by a computer, which simulates the many possible outcomes of an investment decision. During simulation the computer chooses values at random from the probability distribution of each factor affecting the future cash flows. The computer then uses these random values to calculate the return over the project's life, for example.

Then it repeats these calculations a number of times, each time choosing another set of values at random and generating values of the discounted cash flow return. In Figure 2.5 below is the summary of the described simulation process.

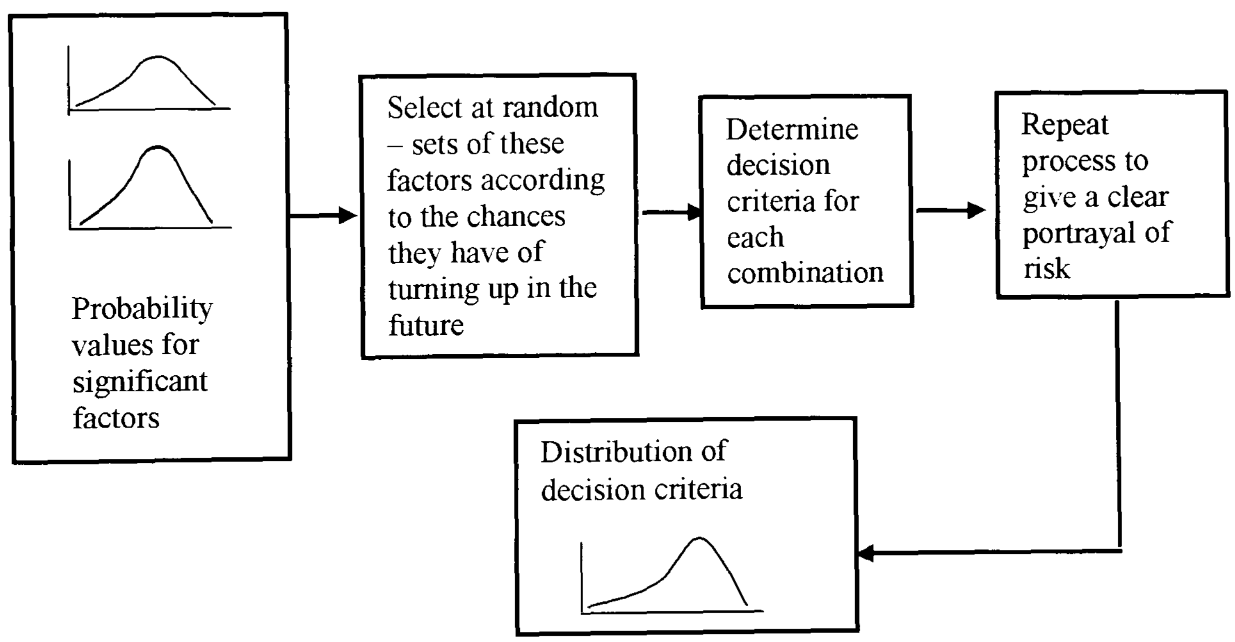


Figure 2.5 Simulation process

(Source: Hertz, 1964)

2.9.3. Criticisms of using simulation analysis

Large number of trials required

One of the criticisms raised in Cooper and Chapman (1987) and Vose (1996) against the used of simulation is that it often requires large number of trials to reduce sampling errors to an acceptable level, and this criticism becomes even more salient if extensive sensitivity analyses are required.

For example, suppose it is forecast that the maximum daily production rate from a certain oilfield is equally likely to have any value between 12 million barrels per day (mb/d) and 20 (mb/d). Then the variable X, say, is Uniformly distributed so that $x \sim U[12,20)$. Its expected value is 16 mb/d. Its standard deviation (s.d.) is " $c / \sqrt{12}$) " where c = largest value - smallest value, = 20 - 12, = 8 mb/d. Hence if, say, 1000 values of the production rates are generated in the simulation the mean value and standard deviation of these 1000 production rates should be extremely close to 16 mb/d and 2.3094 mb/d respectively, or else needless bias will have been incorporated into the process.

Early on in this research it was recognised that a fundamental requirement was to remove any unnecessary variation or randomness from the eventual risk analysis model. Unfortunately for many years the practice here would simply have been to use traditional Monte Carlo sampling methods, where the pure randomness associated with such a method often results in sample outcomes which either exhibit sizeable bias (e.g. a disproportionate number may be very small in value) or require very lengthy and expensive computer runs in order to ensure that the observed frequencies of the simulated values or ranges of values generated from the individual distributions are at least moderately close to the corresponding expected values.

Techniques aimed at reducing the quantity of effort (and thus cost) in order to be representative at a certain level are called Variance Reduction techniques. Although various techniques of variance reduction have been developed over the years to lessen the effort required to achieve specific levels of accuracy (Wilson, 1984; Avramidis and Wilson, 1996), it was the advent of Latin hypercube sampling (Lhs) which revolutionised matters here, since this uses a form of 'stratified sampling without replacement' (Iman and Conover, 1980) to ensure that a sample simulated from a particular population has, for example, values of its sample mean and standard deviation which are very close indeed to the corresponding population values. Thus, Lhs removes a major source of inaccuracy and thus risk.

Much of the work on developing ever more accurate modelling of samples to ensure good sample means and s.d.s ensue can be traced back to the work of Pearson and Tukey (1965).

Today it has became quite easy to develop routines to generate highly accurate and thus representative samples from given individual distributions. This research has achieved this so that, for example, extremely accurate samples from populations with Beta, Uniform, Triangular or Normal distributions can be generated, and it is easy to extend these approaches to deal with many other continuous 'named' distributions, or with distributions defined on the basis of historical relative frequencies.

How Lhs is formalised, the advantages and disadvantages of using it and the comparison of the output results from using Monte Carlo sampling and Latin Hypercube sampling will be discussed in detail, starting in the next chapter, Chapter 3.

Simulation is a powerful method of investigating the riskiness of the projects. However, it still seems to be less favoured by corporate decision makers than is sensitivity analysis using deterministic data. According to Gapenski (1990) and Dingle (1997) the unpopularity of using simulation in the project investment appraisal process perhaps arises because of top management's widespread unfamiliarity with probability concepts, or even fear of the unknown.

According to Newendorp (1985), although simulation has been used for many years, there is beginning to emerge evidence suggesting that, in general, decision makers tend to make their distributions too narrow. That is, when asked to assign a range and distribution to a variable, the tendency is to make the range between minimum and maximum values too small. He clarified the above problem by claiming that this suggests that it is quite difficult to convert from thinking deterministically to thinking probabilistically (as required with simulation). However, it should not affect the choice of using simulation by the decision maker because as far as simulation is concerned the input distributions can have ranges of any magnitude.

Is it always sensible to use simulation? For example, in petroleum economic evaluation the decision maker may feel that simulation only has value in exploration areas where there is sufficient data available upon which to base the distributions. In virgin new exploration areas where there are no data available there is no guaranteed way to define the distributions – and hence, at first sight, no way to make a simulation analysis.

However, Newendorp (1985) argued that quite the opposite is desirable. It is in these new areas where the uncertainty is the greatest where it is most important to consider variability.

The decision maker may only be able to determine or estimate a range of possible values, without knowing the exact shape of the distribution between the limits. But it is essential to consider variability in these new areas, and simulation is the only way this can be done.

For example, historical evidence from broadly similar projects can be used to help choose the distributions. It is likely that the minimum, most likely and maximum values of the individual parameters can be identified, and at worst a Triangular distribution or at best a Beta distribution can be used to model the variables.

Rosenhead's recommendation (1989) that one 'should not pretake the future' has to be ignored here or uses of, and massive benefits arising from, the powerful simulation methodology would almost cease to exist.

Ignoring dependency between variables is a crucial problem when using simulation in risk analysis models. It is also the main question that this research is trying to solve. An example of the problem is as follows.

The expected profitability of an oil field can be expressed as the net present value of a cash flow of gross revenues net of operating costs, capital costs and all tax payments. These revenues and costs will in turn depend on other factors which are subject to economic, physical or political risk, such as possible changes in tax rates, the predicted price of oil per barrel (which will usually be expressed in U.S. dollars per barrel), the dollar-pound exchange rate, inflation (which impacts on the value of the writing down of capitalised costs), the amount of recoverable heavy-end gas (which could, for example, be converted into saleable propane), and so on.

Expert advice might be able to suggest probability distributions for these various factors, together with further partial dependencies between the factors. For example, it would seem reasonable to suppose that the reserves of oil and gas should be positively correlated. Similarly, the greater the gas content is so the greater will be the pressure within the petroleum reservoir, and therefore the greater will be the oil production rate.

On the other hand a British production company would wish to produce high volumes of oil when the price in £ per barrel is high. Hence, for any given price of oil in $/b, if the $ to £ exchange rate increases the company will receive less £s per barrel, so that production would be cut back until more favourable exchange rates return. This indicates that there might be a negative correlation between the $ to £ exchange rate and the oil production rate.

A further factor is that the planners must assume that the production from the field would be halted, perhaps permanently, once the remaining levels of petroleum have dwindled to levels which can no longer support production rates which will generate sufficient revenues to at least offset costs. Consequently, if the price of oil in $ drops then so will the volume of recoverable reserves, suggesting yet another positive correlation between two factors.

Over the past few years there have been many publications on the use of simulation methods for analysing risk. Most explain how to describe a distribution for each random variable and then sample a value from each distribution for each pass using a random number as the entry point in a cumulative frequency distribution of the variables. Many analysts fail to realise that this procedure implies that each random variable is independent of all others (Newendorp and Root, 1976).

In reality, certain important random variables in large projects are dependent, and a realistic appraisal of risk and uncertainty must recognise such dependency relationships, as in the example of oil recoverable reserves and oil maximum production rate above.

One caution to be noted by decision makers is identified in Chapman and Ward (1997), which suggested that simulation makes it relatively straightforward to add large numbers of risks together in a single operation to assess their overall impact. Unfortunately, this convenience can seduce analysts into a naive approach to risk combinations which tends to overlook the importance of dependency between individual sources of risk. It also encourages decision makers to set up the combination calculations to present the end result and to ignore intermediate stages for specification and computational convenience.

In addition, Wall (1997) reviewed that in recent high profile publications the presentation of Monte Carlo simulation-based cost analysis overplays the importance of the choice of which distribution to use to represent input variables and underplays the importance of assessing and including correlations between the variables.

Effects of with-correlation and without-correlation

Wall (1997) showed that correlations between variables of a cash flow model in a Monte Carlo simulation must be recognised explicitly. He carried out two simulation tests based on a complete (216 buildings) data set using Lognormal and Beta distributions, all simulations having 10,000 iterations. His results showed that although the mean of without-correlation simulations is not significantly different from with-correlation simulations there is a significant and substantial difference in the standard deviation.

The most important result is shown by the standard deviation statistic: not including correlations in the simulation can cause a serious mis-estimation of the spread of the distribution. This result was supported by the test statistics.

The standard deviation of the with-correlations distribution is well within the confidence interval of the standard deviation of the observed distribution at the 95% confidence interval, whereas the standard deviation of the without-correlations distribution is well outside the 95% confidence interval. Thus the standard deviation of the with-correlations distribution is not significantly different from the observed, whereas the without-correlation distribution is significantly different from the observed distribution.

The common practice of excluding correlations results in mis-assessment of the risk of the project, which is serious since a key requirement of risk analysis is to assess risk. This result shows that the correlation between risk factors is important when simulation is used to produce estimates of the cash flow model. Neglecting correlations is an erroneous decision and it may well produce misleading results.

Effects of choosing a different distribution

The effect of choosing a different distribution can be observed from the same test which was carried out to show the significant of correlation in simulation by Wall (1997). In the test, Lognormal and Beta distributions were used to observe the significance of the choice of distribution in the simulation model. The test statistics showed that for both Lognormal and Beta distribution, they are broadly similar when compared with the dissimilarity between the without-correlation results and the test statistics.

This further verifies that the importance of calculating and incorporating correlation into simulations outweighs the importance of deciding which distributional form to use to represent the variables of the simulation.

The above criticisms and discussions are further stressed in Uher (1996), who say that the correct assessment of dependence or correlation among risk variables in a simulation model is far more critical in terms of the accuracy of the result than the choice of the probability distributions for the input data.

However, the mechanics of how individual distributions are combined is not always transparent to the decision maker. Together these factors can lead to a failure to appreciate insights from considering intermediate stages of the combination process and dependencies between individual sources of risk. As a consequence, the principal weakness of many simulation analyses, i.e. ignoring the interdependency between variables, is a serious risk in itself.

To overcome the above problem relating to random variable dependencies, one immediate problem would then be how to modify the normal sampling procedures on each simulation to account for observed partial dependencies between random variables

2.10 Key aspects of correlation

2.10.1 Basic definitions

The regression of a variable on one or more other variables provides an indication of the way in which the first variable varies with the second or others, whereas correlation is a measure of how strongly these two or more variables are related to each other, if at all.

The two principal measures of correlation in current usage are Pearson's Product-Moment correlation coefficient and Spearman's Rank correlation coefficient. The former is applicable only when the variables are measured on cardinal scales, and the latter is most appropriately used when the data values are ordinal (Curwin and Slater, 2002).

However, the latter is also often used as a convenient or necessary approximation to the product–moment correlation coefficient, particularly when (as with risk analysis packages such as @RISK or with spreadsheet programmes such as Microsoft Excel) product-moment correlation generating functions are not supplied.

If two variables X and Y are assigned n paired values x_i and y_i in a simulation, for i = 1 to n, then the definition of Pearson's Product-Moment correlation coefficient for this sample is

r_P = covariance(x,y) ÷ [(standard deviation of x) * (standard deviation of y)].

Consequently

$$r_P = (1/n) \sum_{i=1}^{n} (x_i-m_x)(y_i-m_y) \div [\{ \sqrt{(1/n)} \sum_{i=1}^{n} (x_i- m_x)^2 \} * \{ \sqrt{(1/n)} \sum_{i=1}^{n} (y_i-m_y)^2 \}],$$

where m_x and m_y are the sample means of the n values of X and Y respectively.

Alternatively if the values x_i, for i = 1 to n, represent the ranks of the n values of X (where the smallest rank, 1, is assigned to the smallest value) and the y_i values are correspondingly the ranks of the n values of the variable Y, the formula for Spearman's Rank correlation coefficient is easily derived from the above formula for r_P, and is as follows (subject only to a small correction factor if any of the ranks are tied):

$$r_S = 1 - 6 \sum_{i=1}^{n} d_i^2 /(n(n^2-1)), \text{ where } d_i = (x_i - y_i)$$

In each case it can be proved that the least and greatest possible values of r_P or r_S are –1 and 1, and a number of other standard properties of these two coefficients are discussed in Curwin and Slater. For example, r_S is not as sensitive to changes in outlying values as r_P.

Example

Suppose n = 4, and the 4 paired values (x_i,y_i) are as follows:

i:	1	2	3	4
x_i:	3.7	8.6	7.6	8.1
y_i:	2.9	2.0	3.8	3.3

Here the sample means are $m_x = 7.0$ and $m_y = 3.0$. $\therefore$ The covariance of x and y is –0.1150, and the two standard deviations are $s_x = 1.9378$ and $s_y = 0.6595$, so that $r_P = -0.0900$ to 4 d.p.s. The ranks of x_i are 1, 4, 2 and 3 respectively, and the corresponding ranks of y_i are 2, 1, 4 and 3. The d_i values are –1, 3, -2, and 0 (so that the total difference is zero, as always), and the d_i^2 values are 1, 9, 4 and 0. Hence $r_S = 1 - 6 * 14 / (4 * 15), = -0.4$.

A correlation matrix is a symmetric and positive semi-definite matrix (Marsaglia and Olkin (1984)), and is used generally when more than two variables are being simulated. However, even in the direct correlated case of only two variables x and y, say, whose correlation is specified to be ρ, a correlation matrix M may be defined.

Expressed as a table this matrix is:

	Correlation Matrix	
	x	y
x	1	ρ
y	ρ	1

Note that this matrix is symmetric, and that the diagonal elements are both 1. Since it is to be positive semi-definite, its principal minors must both be ≥ 0, so that $1 - \rho^2 \geq 0$, which implies that $-1 \leq \rho \leq 1$.

Although it is not important in the direct context of this research, it is now shown that the user-defined values of partial correlation coefficients when there are more than two variables will have values which are not just individually constrained to being at most 1 in magnitude.

Example

Suppose there are three variables, X, Y and Z, and that the correlations of X with Y and Z are desired to be 0.8 and –0.5 respectively, whereas the correlation between Y and Z is unknown, and equal to ρ, say.

The correlation matrix in tabular form is:

	Correlation Matrix		
	x	y	z
x	1	0.8	-0.5
y	0.8	1	ρ
z	-0.5	ρ	1

The 2x2 minor again yields $-1 \leq \rho \leq 1$. The 3x3 minor is $1 - \rho^2 - 0.8\rho - 0.89$. This is ≥ 0 if $\rho^2 + 0.8\rho - 0.11 \leq 0$, so that $-0.9196 \leq \rho \leq 0.1196$ to 4 d.p.s. It would be erroneous to claim either that ρ is less than –0.9196, such as $\rho = -0.95$, or greater than 0.1196.

Within the remaining chapters of this thesis only two variables will be correlated at any one time, so that it is not necessary to use the concept of a correlation matrix. Note also that the emphasis in this research is on modelling product-moment correlation coefficients, so that after this section little more will be written about properties of rank correlation.

2.10.2 Literature review

Lurie and Goldberg (1998) report that an important requirement relating to the mathematical consistency of a correlation matrix is that it must be positive semi-definite. Their paper deals with correlation in general and so, in this sense, it is equally applicable to product moment or rank correlation matrices.

They describe the circumstances under which a user-defined correlation matrix may not be semi-positive definite, and use a combination of Cholesky decomposition of matrices and Gauss-Newton iterations to generate a revised correlation which is positive semi-definite and "as close as possible to the user's original matrix". A good example is illustrated in Price (2002)

A useful extension of their procedure is derived from the consideration that there may be more certainty about the values of certain of the correlation coefficients than others, so that weights reflecting the individual levels of certainty can be incorporated within the objective function.

This work is carried out in the context of desiring to generate random numbers from a selection of univariate distributions. Thus, the joint distribution of these variables is (except in quite limited circumstances) unlikely to be known or to be able to be predicated, whereas the marginal distributions of the individual variables may be specified by the user, together with specific information on the partial correlations (or possibly on the product moments up to some finite order). They report on the approaches used by other researchers in this field and discuss how problems are recognised to arise in the modelling of correlations if any of the variables have finite bounds (such as the beta or triangular distributions), for example in the work carried out by Vale and Maurelli (1983).

Certain researchers have overcome this difficulty (for example Li and Hammond (1975)), albeit at the cost of great computer computational effort, and yet even then the resulting correlation matrix may not be positive semi-definite.

Clemen and Reilly (1999) examined the problem of constructing a probabilistic model in the contexts of decision analysis and risk analysis.

They explained that typically this is achieved by defining a joint distribution of all the variables as a product of marginal and conditional distributions for the individual random variables.

As the number of desirable variables grows within the model, the required number of probability assessments can grow exponentially. They devised an alternative approach using a copula, together with measures of pair-wise rank correlations. A copula, formally, is a means of expressing a joint cumulative distribution function of a set of random variables as a single function of these variables' marginal cumulative functions, so that no recourse to conditional distributions need be made. Thus the success of the Clemen and Reilly procedure relies on the fact that rank correlations do not depend on marginal distributions.

In this context the copula approach would be equally applicable if the matrix of dependence measures is expressed in terms of Kendall's τ, rather than Spearman's ρ_S.

Unfortunately, in the context of this thesis, this method cannot be applied when it is desired to express the relationships between the variables in terms of the Pearson product-moment correlation, ρ_P, since ρ_P depends on the marginal distributions.

However, this paper is instructive on ways of helping the user or modeller to think about the relationships among the random variables. The paper also has a very clearly developed example to illustrate the procedure. They conclude by raising a key question: "Rather than asking if experts can assess correlations accurately, perhaps we should ask whether they can assess correlations well enough to be useful in the modeling process. The results reported in this paper suggest an affirmative answer." Here, then, clearly, they are in agreement with the observations noted earlier by Wall (1997) and Uher (1996).

Iman and Conover (1982) developed a procedure for generating a desired rank correlation matrix on a multivariate input random variable in a simulation study.

It must be stressed here that the measure of correlation is rank correlation and not product-moment correlation, so that it is relatively unhelpful in the context of this thesis. Because rank correlations are being used, the method is distribution free so that the exact forms of the marginal distributions of the input variables are preserved. The method is quite straight forward to implement within a bespoke simulation study, and is equally applicable to either Monte Carlo or Latin hypercube sampling approaches.

In practice this approach forms the basis for correlation modelling in commercial risk analysis and simulation programmes such as @RISK and Crystal Ball. In chapter 6 of this thesis it is demonstrated that, in contrast with quite accurate sampling from the marginal distributions, the generated rank correlation coefficients are much less accurate.

Iman and Conover observe in effect that "if the sample rank correlation is not satisfactory to the user, nothing prevents the prospective user from generating several (candidate) matrices of Spearman correlations, and then choosing the matrix that provides the most preferred rank correlation." This rejection method could well be long drawn out if the number of variables, n, is large. In this sense the Lurie-Goldberg use of a weighted objective function seems more attractive but, again, this latter can only be used with rank correlations.

Schmeiser (1999) observes that commercial simulation software provides extensive support in creating the logical model, but relatively limited support for creating the input model.

Most of the effort expounded by various researchers, such as Iman and Conover, and Lurie and Goldberg, has been directed at developing complex input models that cater for rank correlations only. However Schmeiser notes that Cario and Nelson (1997) and Chen (1999) "at least tackle the harder problem of providing the desired Pearson correlation".

The NORTA ("NORmal To Anything") method involves a components-wise transformation of a multivariate normal random vector into a random vector with specified marginal distributions for the individual variables. Ghosh and Henderson (2002) observe that the approach is equally valid for both rank and product-moment correlations, which is a definite step ahead from the Iman- and Conover-derivatives.

Chen (2001) developed a procedure to generate n-dimensional random vectors using the NORTA approach. He stated that n(n-1)/2 non-linear equations need to be solved to ensure that the generated n-D random vector has the specified correlation statistics. This method is computationally more complex yet improves the Cario-Nelson algorithm's speed and robustness.

Indeed, it is the question of robustness that raises the largest doubt about the Cario-Nelson algorithm. Ghosh and Henderson (2002) observed that the NORTA method has been shown to fail for some feasible correlation matrices (i.e. the random vector has the given marginal distributions and the generated matrix is an acceptable approximation to the required correlation matrix). They concluded that this feasibility problem becomes steadily worse with increasing n in general, and actually fails in the vast majority of cases even in as low a dimension as n = 17. However, they propose an augmentation procedure that, initially at least, appears to be retaining an accurate approximation as the dimension increases.

2.10.3 Inspiration

The various increasingly-complex derivatives of the Imam and Conover approach are making progress in this general modelling area, but appear to be making ever greater demands of the modeller.

The Iman and Conover approach only works with rank correlations and, although it can cope with modelling quite large dimensions of vectors, the accuracy of the generated correlation matrix usually leaves a great deal to be desired. The NORTA approaches show some promise when dealing with relatively small numbers of variables (perhaps n less than 15 at best), and does seem capable of addressing the product-moment correlation problem, but otherwise the method crashes with disquieting frequency. It is likely that future NORTA-augmentation procedures (along the lines promised by Ghosh and Henderson) will produce a more robust and accurate simulation input model, but this is likely to be at the expense of ever greater demands on resources.

The enormous effort to specify (and fit) the joint distribution is a huge drawback, particularly where larger numbers of variables are concerned, and this (and other) drawbacks make their use impractical for a model of relatively modest complexity. Aiming for the simpler goal of matching only the marginal distributions (i.e. the claimed distributions of the individual variables) and the correlation matrix, may well capture the essence of the dependence between the components while being able to work with easily implementable methods that work well in higher dimensions (Ghosh and Henderson, 2002). Consequently a quite different approach from the NORTA-derivatives would be attractive, and such a one is defined, specified, developed and tested within the remainder of this thesis.

Summary

The first part of this chapter defined the scope of this research by starting with the meaning of uncertainty and risk. Uncertainty is therefore defined as the result of imperfect knowledge about an event and risk is the result of uncertainty. This has lead to identifying the significance of employing risk analysis within the environment with some real world examples from various application areas.

While data used for evaluation are not always known precisely, these parameters may be the best estimates of experienced personnel or they may be based on a very cursory analysis of minimal data. Consequently, decisions must be made in the face of this uncertainty. As a result, risk associated with parameter estimates must be incorporated in any QRA model.

The second part of this chapter analysed the three main approaches used in risk analysis during the capital budgeting evaluation. Simulation is thought to be able to provide an objective evaluation for a project

The use of simulation is recommended as it can represent uncertainty in terms of continuous probability distributions rather than just a few values, and therefore it provides a better replica of a project's real-world risk/return characteristics than does point analysis or sensitivity analysis. It is especially useful when there is a simultaneous change in many variables. Simulation provides a better framework for analysis in that it is easier to estimate a range of values for the variable rather than one best point estimate. The more the uncertainty in estimating a variable, the greater is the advantage of employing simulation (Nanda and Miller, 1996).

However, QRA using simulation has suffered from several drawbacks that have been pointed out in this chapter. The most serious one is ignoring the dependency between these uncertain variables.

This leads to the research question on how to incorporate the dependency into a QRA model using simulation.

The next chapter will demonstrate the structure of the RCM and the methodology used to formulate it. Some pre-requisite know-how and proof of the relevance of methods used in the RCM will be explained before the actual building of the model is discussed in detail.

Finally, Bennett et al (1970) and Vose (1996) remind us that, in the final analysis, any evaluation technique is only as good as the estimates of its input parameters and must be used in conjunction with sound managerial judgement. These techniques only provide management with information tools to aid in the decision-making process.

Chapter 3: The Methodology for A Model Simulating Product Moment Correlation

3.1 Introduction

Chapter 2 concluded that simulation analysis is recognised as an improved approach over point and scenario analysis when uncertainty from the project needs to be incorporated and the risk from undertaking the project needs to be analysed. Nevertheless, it was identified that there are a number of limitations from using simulation analysis.

The two main problems that this research is trying to solve are:

Problem 1 - Ignoring the dependencies among key factors leads to the risk of generating unrealistic random numbers as the input for subsequent calculations.

Problem 2 – Traditionally in Monte Carlo simulation a large number of trials is required in order to generate an observed frequency distribution which closely approximates to the required probability distribution for either key factors or decision criteria.

This chapter in effect describes the various 'building blocks' that have eventually been chosen to contribute to the RCM, for example sorting and shuffling routines. It also describes the two major approaches to sampling from a single distribution: the basic concepts of Monte Carlo sampling, and the relatively recent and very powerful extension of stratified sampling know as Latin hypercube sampling.

The key objective of this chapter is that the samples which are generated and the correlations between the samples should be very accurate reflections of the populations from which they are drawn, in order to maximise the confidence of the planner that the modelling is as correctly representative as possible.

By the end of this chapter, what is arrived at is a means of generating random numbers from U[0,1) using Latin hypercube sampling, with perfect sample means and variances and for which the sample correlation coefficient is not only a product-moment coefficient but is also very accurate.

It is worth noting that this research restricts its scope to handling two correlated variables. Simulating product moment correlations among three or more variables will become a future extension of this research.

One assumption of the Research Model is that the two variables defined in the model have values which are cardinal, so that Pearson's product moment correlation coefficient should be used. Spearman's rank correlation coefficient is not appropriate since it is best used when the data variables are ordinal.

Various authors have shown how Spearman's Rank Correlation Coefficient can be modelled, for example Newendorp and Root (1976), Wall (1997), and Vose (2000).

This research uses a bottom-up approach to produce the output of two sets of correlated sample values. This will begin with the statement of the assumptions of the probability distributions of the individual variables, together with assumptions about various pair-wise correlations between the variables.

The following is the schemata of this chapter. The building block in each step will be discussed before the formulation of the research model is explained.

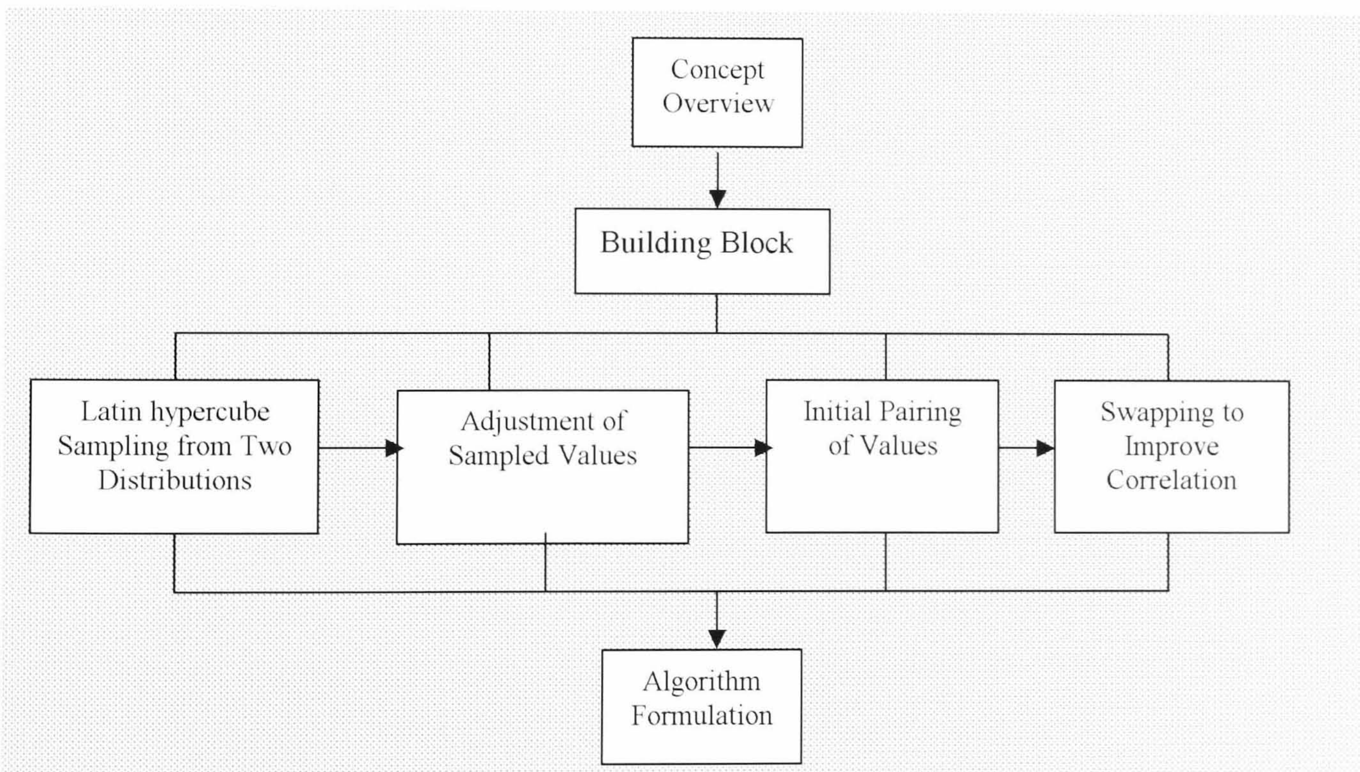


Figure 3.1 Chapter schemata

3.2 An algorithm to generate accurate pair-wise correlations

The two primary objectives here are:

(i) to generate and adjust the two samples in such a way that the two sample means and their standard deviations are equal to the corresponding population parameter values μ_x, σ_x, μ_y and σ_y; and

(ii) to choose a suitable set of pairings of the two sets of sample values without replacement so that the correlation between the two samples is acceptably close to the population product-moment correlation coefficient, ρ.

Suppose the first sample is that of a variable x whose sampled values are x_i for $i = 1$ to n, and for which the expected value and variance are μ_x and σ^2_x. Similarly the second sample consists of n values of a variable y. The key notations used are as follows:

Symbol	**Range or Description**
x_i	Values from the first distribution, for $i = 1$ to n
y_j	Values from the second distribution, for $j = 1$ to n
μ_x and μ_y	The two population expected values
σ_x and σ_y	The populations' standard deviations
ρ or ρ_{xy}	The target product-moment correlation coefficient

The algorithm to achieve these objectives is in two parts, the first of which is described firstly in broad outline in section 3.2.1. and the second in more detail within section 3.3 to 3.6.

Initially, i.e. in Chapter 3, the distributions of x and y is to be restricted to U[0,1), so that $\mu_x = \mu_y = ½$, and $\sigma_x = \sigma_y = \frac{1}{\sqrt{12}}$, but the results are extended to more general distributions in Chapter 5.

3.2.1. Conceptual approach

The conceptual approach is represented in this section to give an overview of the steps related to the RCM, and then the algorithm implemented within the model is explained.

It begins by defining two variables X and Y to have U[0,1) distributions, so that their population means are both equal to ½ and their s.d.s are both equal to $1/\sqrt{12}$. The four major steps are:

<u>Step 1</u>

Generate n representative 'realisations' of X which are sampled randomly in turn from n equi-probable equal-width sub-domains of the overall domain [0,1) of X, and which are adjusted to ensure that the sample mean is ½, and the s.d. is $1/\sqrt{12}$, these n values being stored in the array $\underline{x}$;

<u>Step 2</u>

Generate similarly n representative realisations of Y which are then stored in the array $\underline{y}$;

Step 3

Identify a good starting sequence in which the values from $\underline{x}$ and $\underline{y}$ should be paired in order to achieve a sample correlation which has a first-order approximation to the required correlation ρ; and

Step 4

Adjust the pairings so that the final sample correlation is very close indeed to the required value ρ. The accuracy of the final sample correlation will ultimately (in Chapter 6) be bench marked against that achieved by using the commercial package @RISK.

3.3 Generate a random sample using Latin hypercube sampling

3.3.1. Sampling method used

What is meant by randomness is that the process which produces the number is not deterministic, so that we cannot be sure what number will be produced next.

These random numbers are transformed into samples from the required distribution.

It is often not good enough if the modeller wishes to minimise the sampling errors (the set effect) that occur due to shortish run length (Pidd, 1998). The two most popular sampling methods in current usage are Monte Carlo and Latin hypercube sampling. The latter is a sampling method using descriptive sampling and it is preferred in this research. The advantages of using Latin hypercube sampling method are demonstrated below.

3.3.2. Monte Carlo sampling

Monte Carlo sampling is the oldest, least sophisticated and yet still most popular sampling method used in academia and businesses.

Monte Carlo sampling got its name from the code name of an American project on the atom bomb during the Second World War and not, as some people believe, from the town in Monaco with the same name that is so well known for its casinos (Vose, 1996).

The process of Monte Carlo sampling can be viewed as two steps. First, select a uniformly distributed value between 0 and 1. For example, in Visual Basic, this is a call to the function RND. Second, use the cumulative density function (CDF) for the distribution of its risk factor to identify a value of this random variable.

Monte Carlo sampling satisfies the purist's desire for an unadulterated random sampling method. It is useful if one is trying to get a model to imitate a random sampling from a population or for doing statistical experiments.

However, the randomness of its sampling means that it will over- and under-sample from various parts of the distribution and cannot be relied upon to replicate the input distribution's shape unless a very large number of iterations are performed.

For nearly all quantitative risk analysis modelling, the pure randomness of Monte Carlo sampling is not really relevant. Increasingly users are far more concerned that the model should reproduce the distributions that we have determined for its inputs. Otherwise, what would be the point of expending so much effort on getting these distributions right?

3.3.3. Latin hypercube sampling

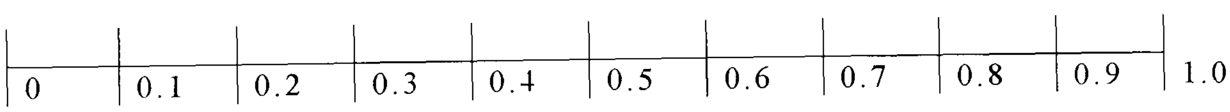
Latin hypercube sampling addresses this issue by providing a sampling method that appears random but that also guarantees to reproduce the input distribution with much greater precision and therefore efficiency than Monte Carlo sampling.

Latin hypercube sampling, or Lhs, is an option that is now available for most commercial risk analysis programmes, for example. @RISK and Crystal Ball. It uses a technique known as 'stratified sampling without replacement' (Iman and Conover, 1980) and proceeds as follows:

To generate n random values from U[0,1) using Latin hypercube sampling:

1) Divide the domain of X into n mutually exhaustive classes: [0,1/n), [1/n,2/n), . . . , [1-1/n,1), so that the probability that X takes a value in any one of these classes is constant at 1/n.

For example, if n = 10



2) These classes are closed at the left and open at the right. E.g. In the first class, [0,1/n], a realisation x_1 will be generated which will satisfy $0 \leq x_1 < 1/n$. The formula used to generate the general value, x_i, is:

$$x(i) \equiv (i-1)/n + (1/n) * \theta, = (i -1+ \theta)/n,$$

for i = 1, 2, . . . , n, where θ is a random number from U[0,1).

Note that an alternative (and inter-changeable) notation for x(i) used in this thesis is x_i.

Example, with n = 10:

i	**θ**	**x(i)**
1	0.2487	0.02487
2	0.1986	0.11986
3	0.9143	0.29143
4	0.8849	0.38849
5	0.4312	0.443312
...	...	...

Thus each class contains exactly one random value. The sample mean, m_x, say, should be approximately ½ and the s.d., s_x, say, should also be approximately equal to $1/\sqrt{12}$. It is illustrated below how this can be achieved when n is an even number.

3.3.4. When n is an even number

In fact if the value of n is restricted to being an even number (such as 10, 20, 100 or 1000, but not 25) it is easy to ensure that the sample mean is exactly equal to ½. To achieve this simply ensure that the n random numbers θ are generated in antithetic pairs which sum to 1.0. For example, if one value of θ is 0.374 then its antithetic value is 1 - 0.374, = 0.626. This is probably the easiest variance reduction procedure to implement.

Example:

θ	$1-\theta$
0.4996	1-0.4996 = 0.5004
0.1433	1 – 0.1433 = 0.8567
0.0019	1 – 0.0019 = 0.9981

Thus 5 random numbers from U[0,1) can generate 2 * 5, = 10, values in a sample of size 10, with each of 5 class of width 1/5 containing two antithetic values. From this can be seen the desirability of making n even, not odd.

Figure 3.2 taken from Vose (1996) illustrates the use of stratification that is produced for 20 iterations of a Normal distribution. It is observed that the intervals get progressively wider towards the tails as the probability density drops away.

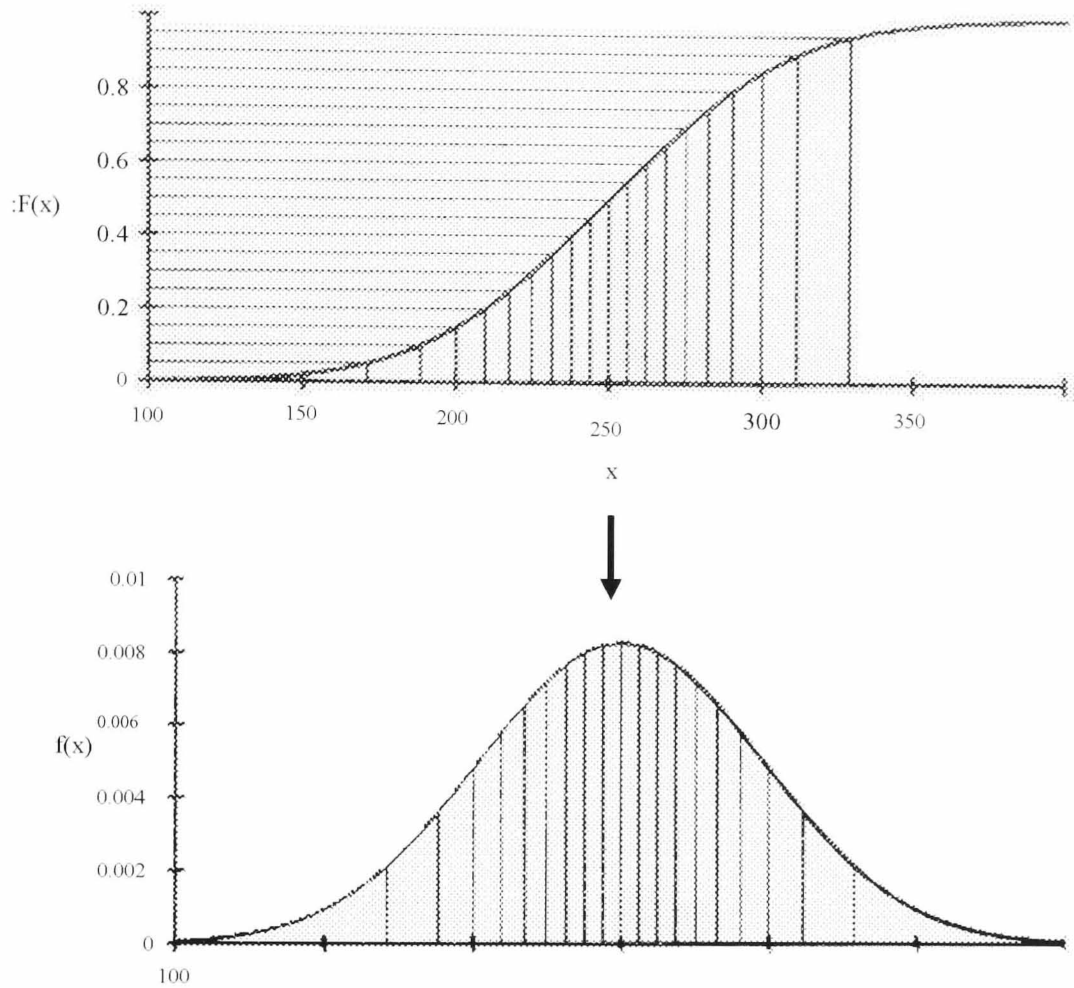


Figure 3.2 Example of the effect of stratification in Latin hypercube sampling

(Source: Vose, 1996)

3.3.5. Comparing results between Monte Carlo and Latin hypercube sampling methods

The improvement offered by Lhs over Monte Carlo can be easily demonstrated. Figure 3.3 taken from Vose (1996) compares the results obtained by sampling from a Triangular(0,10,20) distribution with LHS and Monte Carlo sampling.

The charts of Figure 3.3 show that Lhs consistently produces values for the statistics that are nearer the theoretical values of the input distribution than Monte Carlo sampling. It can also be observed that the histogram of the sample distribution of the 300 iterations resembles a triangle far more closely for Lhs.

Of course, the random nature of Monte Carlo sampling means that another set of simulations might have produced more accurate results were we to have repeated the experiment, but we could never guarantee it.

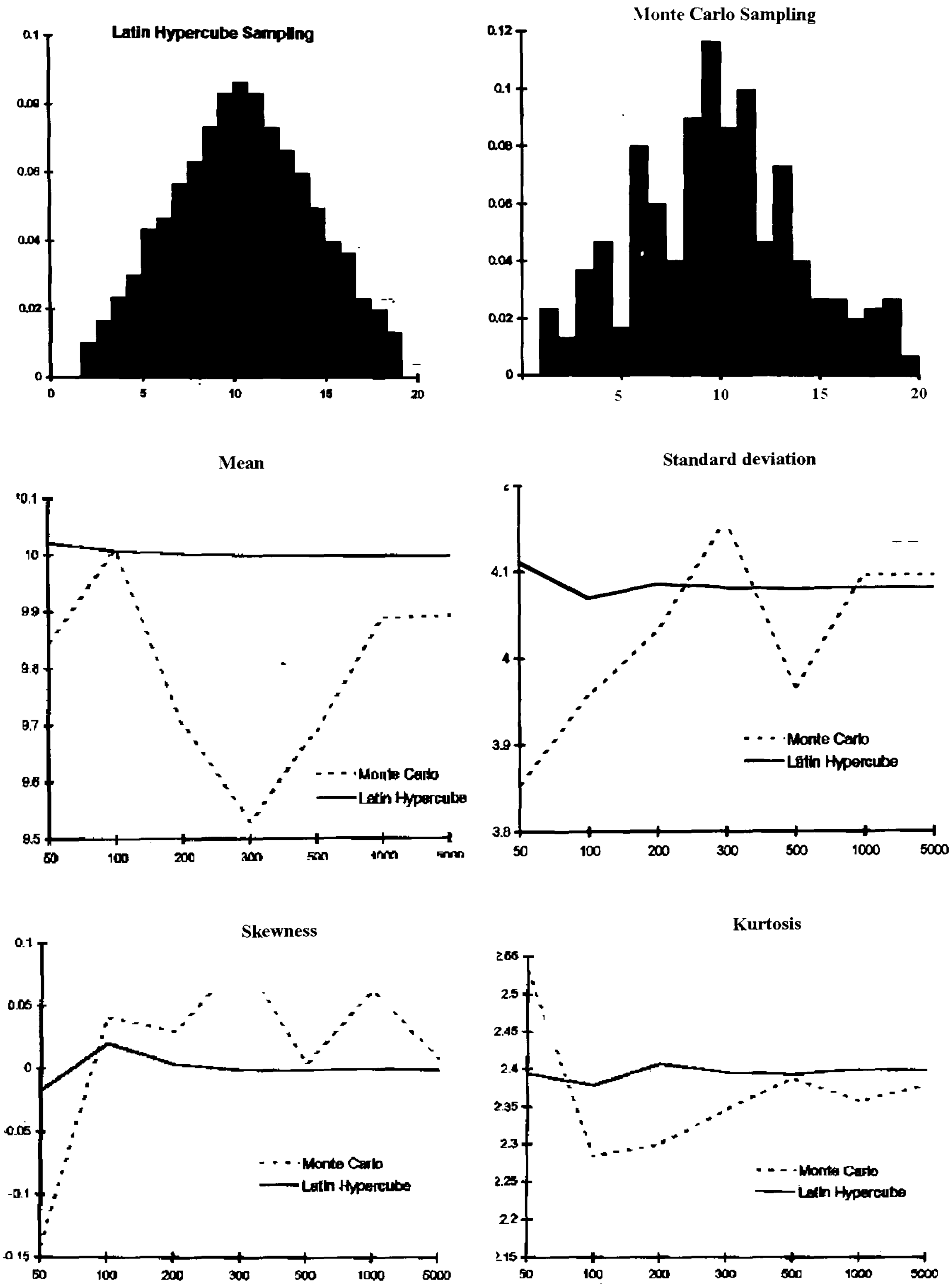


Figure 3.5 Comparison of the performance of Monte Carlo and Latin hypercube sampling (Source: Vose, 1996)

3.4 Maximum absolute errors (MAE) when antithetic variables and Latin hypercube sampling are used

The use of antithetic variables which are used in the Research Model is one of the key methods in the process of variance reduction in discrete event simulation, where variance reduction is aimed at reducing the size of the random sample required to ensure that the sample mean does not differ significantly from the desired value at any specified level of significance. Suppose X is a variable which is distributed as U[0,1), and 2n values are to be simulated, consisting of n pairs of values. In the i^{th} pair let the random numbers (r.n.s) from U[0,1) that are used be θ_i and $1-\theta_i$. This pair of random numbers is antithetic in that their sum is always equal to 1.

This section proves that the maximum absolute errors in the sample s.d. and variance can be reduced when the antithetic method and Latin hypercube sampling is used. Suppose n values of X are sampled from U[0,1) so that the i^{th} value, x(i), has a value generated in the interval [(i-1)/n, i/n). For brevity the analysis is restricted so that n is an even number, although the analysis of the case where n is odd is also easily derived.

3.4.1. Case where antithetic sampling is *not* used.

The greatest sample standard deviation will occur when the value assigned to x(i) either equals the lower class boundary value, (i-1)/n, when $i < n/2$, or equals the upper class boundary, i/n, when $i > n/2$.

On the assumption that the sample mean is adjusted to ½, the sample variance will be:

$$V = \{ [x(1) - 1/2]^2 + [x(2) - 1/2]^2 + \ldots + [x(n) - 1/2]^2 \} / n.$$

Suppose it is expressed in the form that $V = V_{left} + V_{right}$, where the terms contributing to V_{left} will be those for which $i < n/2$. The terms for which $i > n/2$ will constitute V_{right}. It is clear from symmetry that $V_{left} = V_{right}$, and that therefore $V = 2*V_{left}$. Let $n = 2k$.

$$\text{Hence } V = (2/n) * \{ [0/2k - 1/2]^2 + [1/2k - 1/2]^2 + [2/2k - 1/2]^2 + \ldots + [(k-1)/2k - 1/2]^2 \}$$

$$= (2 / n^3) * \{ k^2 + (k-1)^2 + \ldots + 1^2 \}$$

$$= (2 / 8k^3) * \{ k * (k+1) * (2k+1) / 6 \}$$

$$= (1+1/k) * (2+1/k) / 24$$

$$= (1 + 2/n) * (1 + 1/n) / 12.$$

Note that $V \rightarrow 1/12$ as $n \rightarrow \infty$.

The maximum absolute error in the sample variance is thus:

$$(1 + 2/n) * (1 + 1/n) / 12 - 1/12 \quad = [3/n + 2/n^2] / 12.$$

3.4.2. Case where antithetic sampling <u>is</u> used

If the sampling is carried out using antithetic variables then the contributions to the variance from a general linked pair of classes, class 2w-1 and class 2w, will be:

$(x_{2w-1} - ½)^2 + (x_{2w} - ½)^2$

$= ((2w-2 + \theta)/n - ½)^2 + ((2w-1 + 1 - \theta)/n - ½)^2,$

$= (1/4n^2) * \{ (4w - 4 + 2\theta - n)^2 + (4w - 2\theta - n)^2 \}$

$= (1/4n^2) * \{ [(4w-2-n) - 2(1-\theta)]^2 + [(4w-2-n) + 2(1-\theta)]^2 \}$

$= \{ 16w^2 - 8(2+n)w + n^2 + 4n + 8 \} / 2n^2 + 2(1-\theta)^2 / n^2$, for any integer value of n.

The contribution from this pair is clearly maximised when $\theta = 0$.

So, the maximum variance arises in the limiting case in which every first value of the pairs of antithetic variables is at the left hand class boundary (i.e. in the odd numbered classes), and the second sampled value in each pair is at the right hand class boundary (i.e. in the even numbered classes).

This maximum variance is then $(1/2n^3)\Sigma \{16w^2 - 8(2+n)w + n^2 + 4n + 8\}$, summed from w = 1 to n/2

$= [1 + 8/n^2]/12, = 1/12 + 2/(3n^2).$

Hence the maximum error in this variance is $2/(3n^2)$

Note again that $V \to 1/12$ as $n \to \infty$.

E.g. If n = 10, the maximum errors in the variance are "$[3/n + 2/n^2] / 12$",

= 0.32 / 12, = 0.0267 to 4 d.p.s if antithetic sampling is not used, or "$2/(3n^2)$",

= 0.0067 if antithetic sampling *<u>is</u>* used.

Even for such a small sample size, the latter maximum error is approximately a quarter of the former maximum error.

3.4.3. Example of the MAE in the sample mean and variance

It is easy to verify that the maximum absolute error in m_x is half a class width, $= \frac{0.5}{n}$. E.g. If n = 100 then the value of m_x will lie in the range $0.495 \le m_x \le 0.505$. The maximum error would occur when either all values x (i) are at their lower class boundary or, from symmetry, when they are all at their upper class boundary.

The maximum absolute error in the sample variance, s^2_x, will depend on whether the values of θ are generated as antithetic pairs. These maximum errors were derived in section 4.4 above.

If antithetic random number generation is *not* used then the maximum absolute error in the sample variance will be $[\ 3/n + 2/n^2\ ] / 12$. For example, when n is 10 this will be 0.02667, whereas when n = 100 this reduces to 0.00252, and when n = 1000 this reduces further to 0.00025. In practice the error in the sample variance is usually very much smaller than the 'worst case' upper bounds.

If antithetic random number generation *is* used then the formula for the maximum absolute error in the variance is $2 / 3n^2$ so that, if n = 10, the maximum absolute error in the variance is reduced from 0.02667 to 0.00667.

Similarly when n = 100 the maximum error is reduced from 0.00252 to 0.00007, and when n = 1000 it is reduced from 0.00025 to 0.0000007.

Example:

We can compare these maximum absolute errors ("mae") in the variance for several values of the sample size, n:

Sample Size, n	Non-Antithetic Sampling: mae = [3/n + 2/n2] / 12, = N, say	Antithetic Sampling: mae = 2 / 3n2, = A, say	A ÷ N, = 8 / (2 + 3n)
10	0.02667	0.00667	0.25
10	0.00252	0.0000667	0.02649
1000	0.000252	0.000000667	0.002665

Clearly the use of antithetic variables guarantees greater accuracy in the variance, and this accuracy increases with the sample size, n.

Now consider the ranges within which these variances must correspondingly lie.

Example

n = *10*

The population variance is 1/12, = 0.08333 to 5 d.p.s. Using the above expressions for the maximum absolute errors in the variance in the case of samples of size 10, for example, the sample variance V, say, must lie in the range $0.05667 \leq V \leq 0.11000$ if antithetic random number generation is *not* used or in $0.07667 \leq V \leq 0.09000$ if it *is* used. The former is correct only to 1 d.p., whereas the latter is almost correct to 2 d.p.s.

n = *100*

The corresponding ranges if n = 100 are $0.08082 \leq V \leq 0.08585$ and $0.08327 \leq V \leq 0.08340$ respectively, which are correct to (almost) 2 d.p.s and (almost) 4 d.p.s.

n = *1000*

Similarly if n = 1000 the range if antithetic random number generation is not used is $0.08308 \leq V \leq 0.08358$, which is still not certainly correct to 3 d.p.s, but if the antithetic approach is used the range is $0.0833327 \leq V \leq 0.0833340$, which is correct to 5 d.p.s. and is almost guaranteed to be correct to 6 d.p.s.

In summary, the maximum ranges within which the sample variance must lie in these three cases are as follows:

Sample Size, n	**Range of Non-Antithetic Variance**	**Range of Antithetic Varaince**
10	$0.05667 \leq V \leq 0.11000$	$0.07667 \leq V \leq 0.09000$
100	$0.08082 \leq V \leq 0.08585$	$0.08327 \leq V \leq 0.08340$
1000	$0.08308 \leq V \leq 0.08358$	$0.08333207 \leq V \leq 0.08333340$

The population variance of these U[0,1) variables should be 0.083333 to 6 d.p.s, so that this table of the sample variance ranges clearly confirms that antithetic sampling will greatly improve their guaranteed accuracy, and hence greatly increase the confidence in the sample parameter values.

3.4.4. Adjusting to correct the sample mean and variance

The next stage is to linearly transform the values of the x array so that the sample mean of the n values x_1 to x_n is ½ and the s.d. is $1/\sqrt{12}$. This is achieved via the transformation $x(i) \rightarrow ½ + [\ x(i) - m_x\ ] / [\ s_x * \sqrt{12}\ ]$.

This linear transformation is easily verified since the mean value and standard deviation of the transformed x(i) value should be ½ and $1/\sqrt{12}$ respectively:

E[transformed x(i)] = ½ + [½ − ½] / [$s_x * \sqrt{12}$], = ½ − 0, = ½, as required,
and its variance = 0 + Var (sampled value x(i)) / [$s_x * \sqrt{12}\ ]^2 - 0$
= (s_x^2) / 12 (s_x^2), = 1/12, as required.

Similarly transform the n values of the array y.

For example, suppose a sample of 10 values of x(i), 0.00246, 0.19754, 0.24756, etc., have been generated using pairs of antithetic random numbers from U[0,1), and that the standard deviation of these 10 sampled values is s_x = 0.28840 to 5 d.p.s (whereas the population s.d. is $\sigma_x = 1 / \sqrt{12}$, = 0.28868 to 5 d.p.s).

The value of " $s_x * \sqrt{12}$ " is thus equal to 0.99905, and the transform formula yields the three corresponding transformed values of x(i) which, together with the transformed values of the other seven sampled x(i) values, will ensure that the transformed sample mean remains equal to ½, and the sample s.d. becomes $1 / \sqrt{12}$, as required.

i	x(i) before transform	x(i) after transform
1	0.00246	0.00293
2	0.19754	0.19783
3	0.24756	0.24780

Maximum possible change to an individual sampled value from U[0,1)

The maximum possible change to a value can be calculated. For example, if $n = 100$ and antithetic random number generation is used, the absolute change in x(i) will be δ_i, say,

$$= \left| [\, x(i) - \tfrac{1}{2}\,] - [\, x(i) - \tfrac{1}{2}\,] \,/\, [\, s_x * \sqrt{12}\,] \right|$$
$$= \left| [\, x(i) - \tfrac{1}{2}\,] * [\, 1 - 1 \,/\, (s_x * \sqrt{12})\,] \right|.$$

Now, $1 - 1 / (s_x * \sqrt{12})$ will be greatest when s_x is most extreme, for example = 0.28840 as above. Hence the maximum absolute change in $1 - 1 / (s_x * \sqrt{12})$ is 0.00095 to 5 d.p.s, so that no x(i) value will increase or decrease by more than 0.5 * 0.00095, = 0.00048 to 5.d.s., which is 0.0048 of a class width. It should be clear that the largest individual change will be to one of the two extreme values (x_1 or x_{10} in this example). Thus, x_1 above changes by 0.00293 – 0.00246, = 0.00047 (which is just less than the maximum possible change 0.00048 above), whereas the change in x_3 is only 0.00024.

Even so, for the two extreme values x_1 and x_n, a check must be made lest they should become just less than 0 or just greater than 1 respectively, although this is clearly very unlikely. If either of these two extreme-case transforms is infeasible (and this is likely to happen only very infrequently), simply generate a new sample of the variable concerned, transform the values again, and check.

3.4.5. Advantages

Now the values x(1 . . n) form a sample of size n whose:

1) sample mean and s.d. exactly match those of the underlying U[0,1) population;

2) n values are approximately spread out regularly through the range of the cumulative density function or c.d.f.; and thus whose

3) n values are being sampled at approximately equal points throughout the domain of the inverse c.d.f.

Hence if, say, n = 1000 values are sampled and adjusted in this way, and if the observed and expected frequencies are fitted to 10 classes we'd expect a χ^2 goodness of fit test to yield a very small value of the test statistic. Thus the expected frequencies, E_i, would all be 100, and the observed frequencies, O_i, would be close to 100, so that the term "$(O_i - E_i)^2 / E_i$ " would be a small fraction. Hence the value of χ^2 would be small, showing that the sampled data closely mirror the assumed underlying Uniform distribution. A similar process can be found in Avramidis and Wilson (1996).

3.5 Pairing: generating an initial correlation

Now we need to discover which of the y_j values should be paired with each individual x_i value without replacement in such a way that the product-moment correlation of the paired samples is as close as possible to the required correlation coefficient, ρ.

To achieve this, consider the following linear transformation which defines a new array y0(1, . . , n) for values of a, b and c which are to be determined.

$y0(i) = a + b*x(i) + c*\phi(i)$, for $i = 1, 2, \ldots, n$.

The values $\phi(i)$ are to be n separate (and independent) values sampled from U[0,1), and are also to be independent of the x(i) values, so that the covariance of $\underline{x}$ and $\phi = 0$. We require the sample mean and s.d. of the n y0(i) values to be ½ and $1/\sqrt{12}$ respectively, and we require the product-moment correlation between the paired samples to be as close as possible to ρ.

$$\begin{aligned} \text{Now, } E(y0) &= a + b * E(x) + c * E(\phi), \\ &= a + \tfrac{1}{2}b + \tfrac{1}{2}c, = \tfrac{1}{2}, \end{aligned}$$

$$\text{so } 2a + b + c = 1 \qquad (1)$$

$$\begin{aligned} Var(y0) &= E[(y0-E(y0))^2], = E[(y0-\tfrac{1}{2})^2] \\ &= E[(\, b(x-\tfrac{1}{2}) + c(\phi-\tfrac{1}{2})\,)^2], \\ &= b^2 var(x) + c^2 var(\phi), = (b^2+c^2)/12, \\ &= 1/12, \end{aligned}$$

$$\text{So, } b^2 + c^2 = 1 \qquad (2)$$

$\text{Cov}(x,y0) = \rho\sigma_x\sigma_{y0} = \rho/12, = E[(x-½)*(a + b*x + c*\phi - ½)]$

$= E[\, b(x-½)^2 + c(x-½)(\phi-½) \,],$

since the covariance of $\underline{x}$ and $\underline{\theta}$ is zero

$= b/12 + 0, = \rho/12,$

so $b = \rho$. (3)

Hence in (2) $c = \sqrt{(1-\rho^2)}$, and

in (1) $a = ((1 - \rho - \sqrt{(1-\rho^2)}) / 2.$

[Note that equally we could take

$c = -\sqrt{(1-\rho^2)}$, and $a = ((1 - \rho + \sqrt{(1-\rho^2)}) / 2$]

$\therefore y0(i) = [\, 1 - \rho - \sqrt{(1-\rho^2)} \,] / 2 + \rho * x(i) + \sqrt{(1-\rho^2)} * \phi(i),$

for $i = 1, 2, \ldots, n$. (4)

Suppose, for example, that the target correlation is $\rho = 0.30$, and that values of x(i) have been generated for a sample of size 10. Suppose also that values of ϕ(i) have been generated, so that the following table shows the first three pairings of x(i) and ϕ(i), together with the calculated values of y0(i) using formula (4) above.

x(i)	**ϕ(i)**	**y0(i)**
0.00293	0.1470	0.014138
0.19783	0.2578	0.178305
0.24780	0.5013	0.425580

Note that if $\rho = 0$, this reduces to y0(i) = ϕ(i), so that Y0 will be independent of X, whereas if $\rho = 1$, this reduces to y0(i) = x (i), so that Y0 will be perfectly correlated on X.

Similarly if $\rho = -1$, this becomes:

$y0(i) = [\ 1 + 1\ ] / 2 - x(i) + 0, = 1 - x(i)$, which means that these two variables from $U[0,1)$ are then perfectly negatively correlated.

In summary, in these three extreme cases:

ρ	**y0(i)**
0	$\phi(i)$
1	x (i)
-1	1 - x(i)

Otherwise if the magnitude of ρ is closer to 1 than to 0, the term $\rho * x(i)$ will be more influential than $\sqrt{(1-\rho^2)} * \phi(i)$, so that a relatively firm correlation will be created between the values x and y, whereas if the magnitude of ρ is closer to 0 than to 1, the random term $\sqrt{(1-\rho^2)} * \phi(i)$ will be the dominating factor in the generation of y0(i). Algebraically we can verify the following:

$$E[y0] = [\ 1 - \rho - \sqrt{(1-\rho^2)}\ ] / 2 + \rho * E[x] + \sqrt{(1-\rho^2)} * E[\phi]$$
$$= [\ 1 - \rho - \sqrt{(1-\rho^2)}\ ] / 2 + \rho * \tfrac{1}{2} + \sqrt{(1-\rho^2)} * \tfrac{1}{2},$$
$$= \tfrac{1}{2}, \text{ as required.}$$

Hence:

$$\begin{aligned}
Var(y0) &= (1/n)\ \Sigma\ [\ (\ y0(i) - ½\ )^2\ ],\ = (1/n)\ \Sigma\ [\ \{\ \rho * (\ x(i) - ½\ ) + \\
&\quad \sqrt{(1-\rho^2)} * (\ \phi(i) - ½\ )\ \}^2\ ] \\
&= \rho^2 * \Sigma\ [\ (\ x(i) - ½\ )^2\ ] + (1-\rho^2) * \Sigma\ [\ (\ \phi(i) - ½\ )^2] \\
&\quad +2\rho\sqrt{(1-\rho^2)} * \Sigma\ [\ (\ x(i) - ½\ ) * (\ \phi(i) - ½\ )\ ] \\
&= \rho^2 * var(x) + (1-\rho^2) * var(\phi) + 2\rho\sqrt{(1-\rho^2)} * cov(x,\phi) \\
&= \rho^2 * 1/12 + (1-\rho^2) * 1/12 + 2\rho\sqrt{(1-\rho^2)} * 0 \\
&= 1/12, \text{ as required.}
\end{aligned}$$

Similarly the expected correlation between the $\underline{x}$ and $\underline{y0}$ values will be

$$\begin{aligned}
&= (i/n) * \Sigma\ [\ (\ x(i) - 1/2\ ) * (\ y0(i) - 1/2)\ ] \div [\ \sigma_x\sigma_\phi\ ] \\
&= (1/n) * \Sigma\ \{\ [\rho * (\ x(i) - ½\ ) + \sqrt{(1-\rho^2)} * (\ \phi(i) - ½\ )\ ] * [\ x(i) - ½\ ]\} \div [\ \sigma_x\sigma_\phi\ ] \\
&= \{\ \rho * var(x) + \sqrt{(1-\rho^2)} * cov(x,\phi)\ \}\ \div [\ \sigma_x\sigma_\phi\ ],\ =\ \{\ \rho * (1/12) + \sqrt{(1-\rho^2)} * 0\ \} \\
&\quad \div [\ 1/12\ ], \\
&= \rho.
\end{aligned}$$

In practice the sample variance will be close to the required value ρ, as we'll see, but it doesn't necessarily follow from the equation (4) defining $\underline{y0}$ in terms of $\underline{x}$ and $\underline{\phi}$ that Y0 will have a Uniform distribution.

The n values in the $\underline{x}$ array are strictly monotonic increasing, but the values in the $\underline{y0}$ array are probably not, unless the value of ρ is very close to 1.0. The sample mean and s.d. of the generated y0(i) values may be exactly what are required, but the skewness (and kurtosis) of these generated values often differ significantly from the corresponding parameter values of a Uniform distribution.

Hence the array y0 should not be regarded as an ideal sample of values of the variable Y which is ~ U[0,1). Indeed these y0(i) values will be used to decide which individual values of the generated y array will be paired with which of the x(i) values, as will be seen.

3.5.1. The usefulness of the array y0 as a ranking procedure

At this point in the process a sample of n monotonically increasing values of X ~ U{0,1) will have been generated and transformed, so that their sample mean and s.d. are equal to ½ and $1/\sqrt{12}$ respectively. Similarly n monotonically increasing values of Y ~ U[0,1) will have been generated and transformed. These two sets of n values are stored in the arrays x and y respectively.

Also n values of the array y0 will have been generated using ϕ, but these values are almost certainly not monotonically increasing (unless ρ is very close indeed to 1 in value), and the distribution of the variable Y0 need not be U[0,1). These y0 values will be used to decide which individual values in the arrays x and y are to be paired, the supposition being that, since x and y0 have a correlation roughly equal to ρ then, after pairing is completed, so will the arrays x and y. Later (in section 3.6) the swapping routine will be described which should greatly increase the precision of the correlation between x and y, and which will thus go a long way towards achieving the research objectives defined in Chapter 1.

The procedure here is perhaps best developed via an example. The sample size is n = 5, so that antithetic sampling has not been used. The target correlation is $\rho = 0.6$, so that formula (4) simplifies to $y0(j) = -0.2 + 0.6\ x(i) + 0.8\ \phi(i)$. The table of the simulated (and transformed) values of x(i) and y(i) now follows, together with generated values of $\phi(i)$ and hence the calculated values of y0(j).

All the values in these four arrays are written to only 2 d.p.s for readability, but even so the sample means of the x and y arrays are equal to 0.5, and their variances are close to 1/12 (= 0.08333). The sample means and variances of the ϕ and y0 arrays are irrelevant.

The fifth and sixth columns in the table show the ranks of x(i) (which are monotonically increasing) and y0(j) (with the smallest value in each case having rank 1, etc.). The eight and ninth columns show the values and ranks (currently monotonically increasing) of the values y(j), where the index "j" has been chosen deliberately in order to distinguish it from the index i of x(i). The ranks of y(i) are also monotonically increasing.

i	x(i)	ϕ(i)	y0(j)	Rank of x(i)	Rank of y0(j)	j	y(j)	Rank of y(j)
1	0.10	0.72	0.436	1	3	1	0.14	1
2	0.30	0.19	0.132	2	1	2	0.21	2
3	0.46	0.35	0.356	3	2	3	0.50	3
4	0.74	0.56	0.692	4	5	4	0.78	4
5	0.90	0.41	0.668	5	4	5	0.87	5

The sample mean and variance of the x(i) values are 0.5 exactly and 0.0838 respectively, so that the sample mean is perfect and the variance, which should be 0.0833, is quite acceptable for values recorded to 2 d.p.s only.

The sample mean and variance of the y(i) values are 0.5 exactly and 0.0858.

Now re-order the y(j) values so that their ranks are the same as those of y0(i).

For example the first value in y will be the third ranked, which is 0.50, and the second will be the first ranked: 0.14. The resulting ranks of the re-ordered values in y will in effect be the inverses of the ranks of y0.

The table of values of x(i) and the correspondingly paired y(i) values will thus be as follows:

i	x(i)	Rank of x(i)	Re-ordered y(i)	Rank of y(i)
1	0.10	1	0.50	3
2	0.30	2	0.14	1
3	0.46	3	0.21	1
4	0.74	4	0.87	4
5	0.90	5	0.78	5

The sample product-moment correlation coefficient of these five paired values is easily calculated to be 0.646 to 3 d.p.s, which for a first attempt (and with such a small sample size) is a creditable first approximation to the target value of 0.6.

The procedure covered in the following section will usually enable the precision of the set of pairings in this first approximation to be greatly improved.

3.6 Swapping

This is to be the key step of this algorithm: identifying how this approximation can be improved. Figure 4.2 summarises the approach that is described in this section.

Now, the sample correlation, r_{xy}, is equal to the sample covariance divided by the product of the two sample standard deviations, so that

$\Sigma (x_i - ½) * (y_i - ½)$

$= n * r_{xy} * (1/\sqrt{12})^2,$

$= \Sigma x_iy_i - ½\Sigma (y_i - ½) - ½\Sigma (x_i - ½) - \Sigma ½*½,$

$= \Sigma xy - n/4.$

Ideally this Σxy should be equal to $n * \rho * (1/\sqrt{12})^2 - n/4$, so that the error in Σxy is $n * (r_{xy} - \rho) / 12 + n/4 - n/4, = n * (r_{xy} - \rho) / 12$.

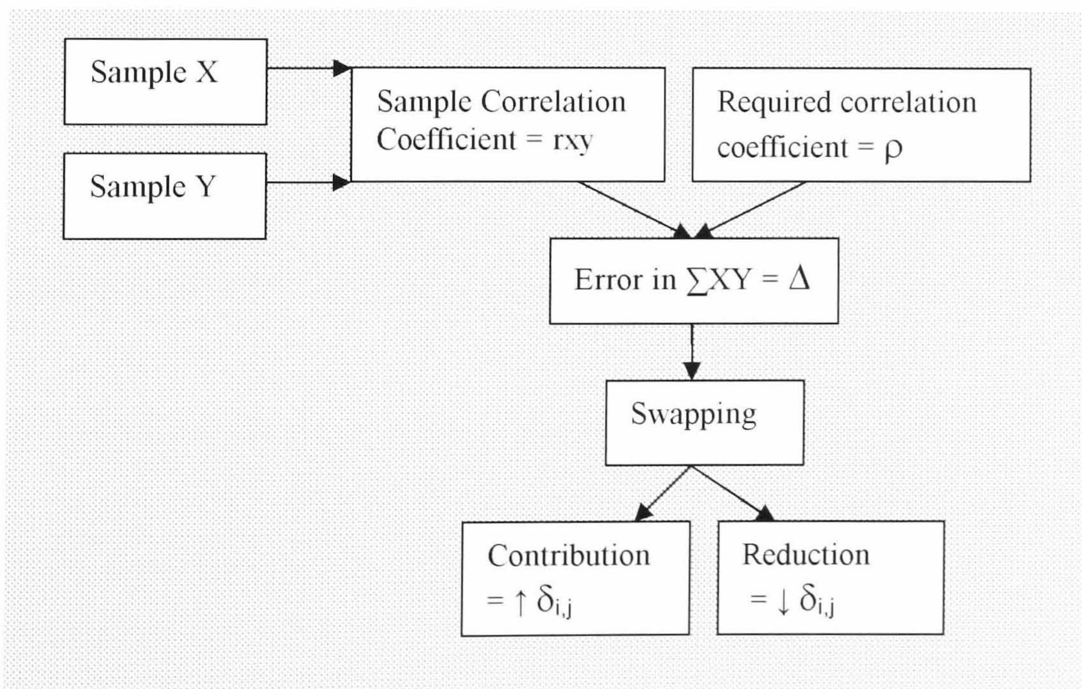


Figure 3.4 Swapping approach

Hence we would like to reduce Σxy by (n/12) * (r_{xy} - ρ), = Δ, say, which could be positive or negative (of course, if it is zero then the sample correlation is already perfect!).

Suppose two of the pairings contributing to Σxy are (x_i,y_i) and (x_j,y_j), where $j > i$, so that their contribution to Σxy is $x_iy_i + x_jy_j$, and consider the effect of swapping the order of these two y values, so that the two pairings would become (x_i,y_j) and (x_j,y_i).

The sample means and s.d.s would be unchanged, but the contribution to Σxy would become $x_iy_j + x_jy_i$, and the reduction in Σxy would be $x_iy_i + x_jy_j - x_iy_j - x_jy_i$, $= (x_j - x_i) * (y_j - y_i)$, $= \delta_{i,j}$, say. Similarly if $\Delta < 0$ then define $\delta_{i,j}$ to be the required increase in Σxy: $\delta_{i,j} = (x_j - x_i) * (y_i - y_j)$.

Hence if $0 < \delta_{ij} < 2|\Delta|$ then swapping the order of y_i and y_j within the array y will improve the required accuracy of the sample correlation.

Hence scan each combination of the values of x_i, y_i, x_j and y_j, and compute the test statistic δ_{ij}. Find the optimal value of i and j (if any) which will then make the most progress in transforming the value of r_{xy} to ρ, with reference to the appropriate double inequality statement bulleted above. This means that scanning of the pairs of values has to be over the ranges $1 \leq i < n$ and $i < j \leq n$, so that the number of combinations of i and j will be ½n(n-1).

Note that, since $x_j - x_i > 0$, if $\Delta > 0$ so that the sample correlation is to be decreased, we only have to consider values y_i and y_j for which $y_j > y_i$. There are similar economics if $\Delta < 0$.

Thus, if values of i and j can be found which will improve the sample correlation in this way, swap the two values of y, and repeat the process, continuing until no further progress can be made. The experience so far is that in the great majority of cases between 1 and 3 iterations are necessary.

The above algorithm is demonstrated in the next chapter by using an example where n is small, = 10, and the target sample correlation is 0.7. Steps 1 to 3 of the algorithm generate samples with correlation 0.6051, and one iteration of step 4 increased the correlation coefficient to 0.6931, which has over 99 percent accuracy. No further improvements were found. Put another way, the reduction in the error exceeds 98.8%, which is highly satisfactory.

Summary

This chapter has demonstrated the method for simulating accurate correlations between two samples drawn from U[0,1) variables, in which the sample mean values and standard deviations are equal to their corresponding population parameters.

This is taken a step further in Chapter 4 when a larger example is worked through. This chapter also examines how the various building blocks are modelled.

In Chapter 5 the algorithm developed in the current chapter will be extended to deal with variables having more general distributions, such as beta or normal. In Chapter 6 the results of a large number of tests of the RCM are reported, and these results are compared with the corresponding results from an industry-standard risk analysis package.

Chapter 4: Simulating Accurate Correlations Between Two Samples: A Full Example, and the Specification of the RCM

4.1 Introduction

This chapter is in two parts. The main objective of the first part is to demonstrate the formulation and building blocks discussed in Chapter 3 with a fully detailed step by step example. Here the two variables still have U[0,1) distributions.

The RCM is a computer based model, so that in the second part of this chapter a flow chart and sequence diagram demonstrate the construction of the RCM. The function of each process, the input it requires, and the output it generates must be specified clearly, so that the flow of data through the model can be clearly identified and understood.

4.2 An illustrative example of the algorithm

Example[1]: Generate two samples of size 10 from populations which are U[0,1) and whose product-moment correlation coefficient is 0.7.

Step 1 – generate X

Divide the domain of X, [0,1), into 10 mutually exhaustive classes:
= [0.0,0.1), [0.1,0.2), [0.2,0.3), and so on, up to [0.9,1.0).

Next generate the 10 random numbers, θ, say, from U[0,1) to be stored in array $\theta(1..10)$[2]. Choose the 10 values of θ to consist of 5 antithetic pairs, so sample 5 values of θ i.e. : $\theta_1, \theta_3, \theta_5, \theta_7, \theta_9$.

$\theta_1 = 0.02462$	$\theta_2 = 0.97538$
$\theta_3 = 0.47559$	$\theta_4 = 0.52441$
$\theta_5 = 0.95135$	$\theta_6 = 0.04865$
$\theta_7 = 0.31880$	$\theta_8 = 0.68120$
$\theta_9 = 0.76934$	$\theta_{10} = 0.23066$

[1] This example, in which the very small sample size is 10, is not particularly practical from the simulation point of view. However it means that the steps of the algorithm can be viewed without running the risk of being lost in masses of data.

[2] Note that, if required, the 10 values in θ could be shuffled into a randomised order (as illustrated in Chapter 3).

Now generate 10 values of x0, one in each of these classes, using the general formula:

$x0(i) \equiv (i-1)/n + (1/n)*\theta(i), = (i - 1 + \theta(i)) / n$, for $i = 1, 2, \ldots, n$, where $n = 10$.

For example, in the third class, x0(3)
$= 0.2 + \theta(3) * 0.1$,
$= 0.2 + \theta_3 * 0.1, = 0.2 + 0.47559 * 0.1, = 0.24756$.

The 10 values of x0 will thus be:

0.00246, 0.19754, 0.24756, 0.35244, 0.49514, 0.50486, 0.63188, 0.76812, 0.87693 and 0.92307.

Because antithetic random numbers have been used in the generation of the array x0 the sample mean value will be exactly ½, = m_x, and the variance, V, is calculated to be 0.08349 to 5 d.p.s, whereas the population variance is 0.08333, so that the relative error is only 0.2 percent. Note that this value of V lies within the range $0.07667 \leq V \leq 0.09000$ derived earlier and is much closer to the true value, 0.08333, than the limits give reason to expect.

The sample s.d. = $s_x = 0.28895$.

Now transform the x0 values to become the final x values via
$x(i) = ½ + [x0(i) - mx0] / [sx0 * \sqrt{12}]$.

i.e. $x(i) = ½ + (x0(i) - ½) * 0.99905$, so that the sample mean of the x(i) values is ½ and their s.d. is $1/\sqrt{12}$.

The 10 adjusted values, x(1 . . 10), are as follows (to 5 d.p.s):

i	x0(i) before transform	x(i) after transform
1	0.00246	0.00293
2	0.19754	0.19783
3	0.24756	0.24780
4	0.35244	0.35258
5	0.49514	0.49514
6	0.50486	0.50486
7	0.63188	0.63175
8	0.76812	0.76787
9	0.87693	0.87657
10	0.92307	0.92267

The sample s.d. was fractionally too large (0.28895 versus 0.28868), so that the effect of the multiplier 0.99905 above has been to bring all the values slightly in towards the sample mean, ½. The two middle values are not changed to 5 d.p.s, but the extreme values are changed the most. For example the first value is changed from 0.00246 to 0.00293.

As mentioned earlier, it is possible that a result of such a transformation (or scaling) could be that either x(1) might become smaller than 0 and/or x(n) might exceed 1, in which case a new sample x0 should be generated. In this case this isn't necessary.

Step 2 - generate Y

In the illustration of this step, all calculated values are correct to 4 d.p.s.

Using Latin hypercubes generate 10 values of Y, and store them in the array yLh0, so that yLh0(1) to yLh0(10) will be (in this example) 0.0193, 0.1807, 0.2685, 0.3315, 0.4668, 0.5332, 0.6372, 0.7628, 0.8076, and 0.9924. The sample mean here is automatically equal to ½, and the s.d. is $s_{yL} = 0.2902$. The transform formula (4) then transforms the array yL into y, say, via the formula

$$y(i) = \tfrac{1}{2} + [\, yL(i) - \tfrac{1}{2} \,] / [\, 0.2902 * \sqrt{12} \,]$$

Note that, as with the generation of the 10 values of x0, the random numbers have been generated in antithetic pairs so that, for example, the first two values yL(1) and yL(2) are 0 + 0.1 * 0.193 and 0.1 + 0.1 * (1-0.1933) respectively. i.e. The 5 pairs of antithetic random numbers from U[0,1) have not been shuffled. Next transform these variables into a new array, yLh(1..n), which has its s.d. equal to $1/\sqrt{12}$.

i	Generated value of yLh0(i)	Transformed value: yLh(i)
1	0.0193	0.0218
2	0.1807	0.1824
3	0.2685	0.2698
4	0.3315	0.3323
5	0.4668	0.4669
6	0.5332	0.5331
7	0.6372	0.6365
8	0.7628	0.7614
9	0.8076	0.8060
10	0.9924	0.9898

Step 3 - pairing

Next we must decide which individual values in x are to be paired with which values in yLh.

First we generate the values of the yR0 array, where the i^{th} term is defined to be:

$yR0(i) = [\ 1 - \rho - \sqrt{(1-\rho^2)}\ ] / 2 + \rho * x(i) + \sqrt{(1-\rho^2)} * \phi(i)$, for $i = 1, 2, \ldots, n$.

To achieve this generate values of $\phi0(i)$, which will be 10 r.n.s from U[0,1) in 5 antithetic pairs and located within Latin hypercubes.

For example, the first two values are:

$\phi0(1) = (\ 1 - 1 + 0.6892\ ) / 10, = 0.06892$; and
$\phi0(2) = (\ 2 - 1 + (\ 1 - 0.6892\ )\ ) / 10, = 0.13108$

r.n.s	Antithetic pairs $\phi0(i)$	
0.6892	$\theta_1 = 0.06892$	$\theta_2 = 0.13108$
0.879	$\theta_3 = 0.28790$	$\theta_4 = 0.31210$
0.3003	$\theta_5 = 0.43003$	$\theta_6 = 0.56997$
0.6220	$\theta_7 = 0.66220$	$\theta_8 = 0.73780$
0.8262	$\theta_9 = 0.88262$	$\theta_{10} = 0.91738$

Of course now yR0(i) will generally take small or large values within its class as $\phi0(i)$ is small or large respectively. This could exaggerate the correlation so, having generated the n (= 10) values of $\phi0(i)$ we now shuffle them into a random order to become the values of the array ϕ which will be used in the generating formula above for yR0(i).

These reordered values are, for example,

ϕ = { 0.66220, 0.43003, 0.91738, 0.28790, 0.06892, 0.56997, 0.13108, 0.31210, 0.73780, 0.88262 }.

Thus, for example, $\phi(1) = \phi0(7)$, and $\phi(2) = \phi0(5)$.

Next calculate the values of yR0(i) using the values of x(j) derived in Step 1 and the values of ϕ(j) listed above:

$yR0(i) = [\ 1 - \rho - \sqrt{(1-\rho^2)}\ ] / 2 + \rho * x(i) + \sqrt{(1-\rho^2)} * \phi(i)$, for i = 1, 2, ..., n.

For example, x(1) = 0.00293 and $\phi(1)$ = 0.66220, so that

$yR0(1) = [\ 1 - 0.7 - \sqrt{0.51}\ ] / 2 + 0.7 * 0.00293 + \sqrt{0.51} * 0.66220,$

$= 0.2679$ to 4 d.p.s.

The 10 values of yR0(i) are:

0.2679, 0.2385, 0.6215, 0.2453, 0.1887, 0.5534, 0.3288, 0.5533, 0.9334, 1.0691.

[Within the programme these values are then rearranged in order of increasing size, to become the array yR, in which case they would be

0.1887, 0.2385, 0.2453, 0.2679, 0.3288, 0.5533, 0.5534, 0.6215, 0.9334, 0.9691.]

The key aspect here is, however, that in general the above values of yR0(i) will not be monotonically increasing, so that we need to identify their individual ranks. For example, yR0(1) = 0.2679, and this is the 4th largest of the values in the array yR, so that the rank of yR0(1) = 4. Similarly we can identify the ranks of the other 9 values of yR0(i), as in the following table, Table 4.1.

The sixth column in Table 4.1 contains the values of yLh which were derived above in Step 2, and so these are monotonically increasing with sample mean ½ and variance 1/12.

i	x(i)	yR0(i)	yR(i)	Rank of yR0(i)	yLh(i)	y(i)
1	0.0029	0.2679	0.1887	4	0.0218	0.3323
2	0.1978	0.2385	0.2385	2	0.1824	0.1824
3	0.2478	0.6215	0.2453	8	0.2698	0.7614
4	0.3526	0.2453	0.2679	3	0.3323	0.2698
5	0.4951	0.1887	0.3288	1	0.4669	0.0218
6	0.5049	0.5534	0.5533	7	0.5331	0.6365
7	0.6318	0.3288	0.5534	5	0.6365	0.4669
8	0.7679	0.5533	0.6215	6	0.7614	0.5331
9	0.8766	0.9334	0.9334	9	0.8060	0.8060
10	0.9227	1.0691	1.0691	10	0.9898	0.9898

Table 4.1 Paring process

Note that it has happened here that yR0(10) exceeds 1 (being equal to 1.0691 to 4 d.p.s), but this isn't important since the distribution of yR0 is *not* intended to be U[0,1). The use of the array yR0 is simply to generate the rankings, so that the array of yLh(i) values can be reordered to have the same ranks. This reordered version of yLh is named y.

For example, because the rank of yR0(1) is 4, the value of y(1) will be the 4^{th} largest value in the monotonic increasing array yLh(i): 0.3323. Continue in this way to complete the seventh column in Table 4.1.

The two arrays x and y in the second and seventh columns have their sample means equal to ½, and their variances are 1/12, = 0.0833 to 4 d.p.s. Their covariance should be approximately $\rho*\sigma_x*\sigma_y$, = 0.7 * (1/12), = 0.0583.

However, the generated covariance of the x and y values is 0.0504, not 0.0583, and so their correlation is 0.6051, not 0.7. Hence our goal now is to increase this the sample correlation to a value closer to 0.7 if possible.

Step 4 - swapping

The final step in the algorithm is to swap pairs of values in the array y(1..n) until the correlation coefficient between x and y is as close as required. Figure 4.2 documents the swapping process in this example for cross referencing throughout the demonstration below.

The covariance = (1/n) *[Σ { x(i)*y(i) } - $\mu_x*\mu_y$], summed from i = 1 to n. Write the sum simply as Σxy , so that Σxy *should* take the value
n * [0.0583 + 0.5*0.5], = 3.0833 to 4 d.p.s.

However the actual current value of Σxy is easily calculated to be 2.9771, so that the ideal net increase in Σxy is desired to be 3.0833 – 2.9771, = 0.1062 (with all calculations being carried out to 4 d.p.s.). This can be described as the target increase in Σxy, and in Table 4.2 below these three values are shown in the third, fourth and fifth rows of the column corresponding to each iteration.

Consequently in each iteration the way in which the values in x and y are paired is adjusted so that the current inaccuracy in Σxy is increased (or decreased, as appropriate) as much as possible. This is achieved by sequentially swapping pairs of values in y, y(i) and y(j), say, until no further improvements can be identified.

Seeking always the largest possible change is a "greedy" approach, and is also known as a "steepest descent" approach, these terms being used frequently in other applications of Operational Research.

Now we'll consider the actual changes or "swaps" to make to pairs of y(i) and y(j) values, iteration by iteration, until no further improvements can be identified.

'Swap' Iteration 1

Target: Σxy should be increased by 0.1062, or as close as possible to this value.

i.e. Currently the value of Σxy is 2.9771, but it should be 3.0833 to 4 d.p.s, so that the target *increase* in Σxy is 0.1062 (= 3.0833 – 3.0043). Note that in this example the value of Σxy is to be increased but could, equally well, need to be decreased. In the notation defined earlier the value -0.1062 is assigned to Δ. I.e. Previously Δ was defined to be the required *decrease* in Σxy.

Thus, suppose two of the pairings contributing to Σxy are (x_i,y_i) and (x_j,y_j), where $j > i$, so that their contribution to Σxy is $x_iy_i + x_jy_j$, and consider the effect of swapping the order of these two y values, so that the two pairings would become (x_i,y_j) and (x_j,y_i). The sample means and s.d.s would be unchanged, but the contribution to Σxy would now become $x_iy_j + x_jy_i$, and the *increase* in Σxy would be $\delta i,j = (x_j - x_i) * (y_i - y_j)$. The ten values of y_i and y_j are currently as shown in the second and third columns in Table 4.2.

Now the values in x are strictly monotonic increasing so that $x_j > x_i$ (when $j > i$), and thus $x_j - x_i > 0$. $\therefore$ We need only to consider elements in the y array for which $y_i - y_j$ is > 0 in this example. Hence, for example, taking $i = 1$ and $j = 3$ will not work since $y_i - y_j = y_1 - y_3 = 0.3323 - 0.7614$, and this is *not* > 0.

Now, in this manual calculation, it is simply a case of finding two values of i and j such that $\delta i,j$ is > 0, and then scanning through the values in Table 2 to find better improvements if possible.

Note also that it doesn't matter if the increase in Σxy exceeds $\delta i,j$, as long as the net increase is less than 2 * $\delta i,j$ so that the value of Σxy becomes closer to the target than before.

The first step here is clearly achieved by swapping y_1 and y_2, so that the value of $\delta 1,2$ = (0.1978-0.0029)*(0.3323-0.1824), = 0.0292. This isn't a big improvement in Σxy, but it still *is* an improvement, and so will become the "incumbent" increase. Now we look for other values of y_j for j > 2 to swap with y_1 such that the incumbent increase will be improved, and so on.

Thus keeping i = 1, the best improvement is obtained by setting j = 5, so that the incumbent value of $\delta i,j$ is $\delta_{1,5}$ = (0.4951-0.0029)*(0.3323-0.0218), = 0.1528. Note that this value makes Σxy too big, being 3.1299, but it will still be closer to the target sought, which is 3.0833, the excess now being only 0.0466 compared with the value 0.1062 before the start of this iteration. i.e. The resulting improvement in the correlation coefficient if we stopped here would already be over 50 percent: the value of r_{xy} is now 0.7559.

Now try taking i = 2 and considering j ≥ 3. The best case here is when j = 5, and yields $\delta_{2,5}$ = 0.0477, which gives a shortfall of 0.0585 in the target improvement in Σxy, compared with the excess of 0.0466 reported above, so that the incumbent pairing is still i = 1 and j = 2.

Continue in this way until we discover that the best improvement is obtained when i = 3 and j = 7 , when $\delta_{3,7}$ = (0.6318-0.2478)*(0.7614-0.4669), = 0.1131, so that Σxy is then equal to = 0.1131 + 0.29771, = 3.0902. Then r_{xy} becomes equal to 0.7082, so that over 93.5 percent of the initial error in the value of the product-moment correlation coefficient has been corrected already, in just one iteration.

Hence, as a result of iteration 1, swap y3 and y7, to yield the updated y vector of values in the fourth column of Table 4.2 at the start of Iteration 2.

'Swap' Iteration 2

The current value of Σxy is 3.0902 and should be increased by 3.0833 – 3.0902 = -0.0068 ideally (adjusting for 4 d.p. accuracy), or as close as possible to this value. The negative sign shows that Σxy should actually be decreased.

This time the best improvement to Σxy is -0.0084 when y_9 and y_{10} are swapped, so that Σxy becomes 3.0849, and the sample correlation becomes 0.6980. If any further improvement is possible, Σxy should ideally be increased by 0.0016.

'Swap' Iteration 3

On inspection no further improvement can be made. As a result of these two swaps, the product-moment correlation coefficient has increased from 0.6051 to 0.6980, so that the improvement in the error is almost 97.9 percent.

It is important to observe that this method is a heuristic. Although the results quoted later in this thesis indicate that almost always (if not always) great improvements in the sample correlation can be achieved, an alternative swapping (or other) heuristic might be identified which would achieve even better results.

In conclusion in this example, the two samples both have sample mean equal to ½ and s.d. equal to $1/\sqrt{12}$, and their sample product-moment correlation coefficient is 0.6980, which is only around one quarter of one percent less than the target value, 0.7.

The table recording the targets and achievements in these iterations is as follows:

	ρ = 0.7: **Swapping:**	**Iteration 1**	**Iteration 2**	**Iteration 3**
	Current Sample Σxy →	2.9771	3.0902	3.0817
	Ideal Required Σxy →	3.0833	3.0833	3.0833
	Target Increase in Σxy →	0.1062	-0.0068	0.0016
	Correlation before iteration:	**0.6051**	0.7082	**0.6980**
i	**x(1..10)**	**y(1..10)**	**y(1..10)**	**y(1..10)**
1	0.0029	0.3323	0.3323	0.3323
2	0.1978	0.1824	0.1824	0.1824
3	0.2478	**0.7614**	0.4669	0.4669
4	0.3526	0.2698	0.2698	0.2698
5	0.4951	0.0218	0.0218	0.0218
6	0.5049	0.6365	0.6365	0.6365
7	0.6318	**0.4669**	0.7614	0.7614
8	0.7679	0.5331	0.5331	0.5331
9	0.8766	0.8060	**0.8060**	0.9898
10	0.9227	0.9898	**0.9898**	0.8060
		↓	↓	↓
Values of y(i) to Swap →		y_3 and y_7	y_9 and y_{10}	No more!
Achieved Change in Σxy →		0.1131	-0.0085	-----

Table 4.2 Swapping process

4.3 The computer based RCM

In Chapters 3 an approach was developed for generating a set of correlated pairs of random numbers, each variable having the underlying Uniform distribution U[0,1), and the procedure was demonstrated in section 4.2. The outcome of the example in section 4.2, i.e. the generated product-moment correlation coefficient of the two samples, was tested against the required correlation coefficient and shown to be within a satisfactory range. The algorithm required for achieving this objective is written in the Gen2Corr function in the computer model using Microsoft Visual Basic (VB).

The next chapter, chapter 5, will then apply the formulated algorithm from Chapters 3 and 4 to a selection of more general probability distributions, i.e. the general Uniform, Triangular, Normal and Beta distributions. The process of transforming the two sets of random numbers into these distributions is written in the TwoDist function in the RCM.

Selecting which programming language was to be used in this research was not a significant issue. The computer model developed in this research is used as a tool for testing and verifying if the algorithm designed and formulated in this research is achieving its objectives, namely accurately modelling the Pearson correlation between two continuous variables in the simulation process.

This section pictures the formulation of the RCM via flow chart and sequence diagrams. Each process will be explained briefly.

4.3.1. Specifying the RCM – flow chart

Within the implementation stage, the designed model can be presented in a more apparent manner, firstly, by a flowchart as Figure 4.1, which is a graphic representation of the steps in the solution of a problem, in which symbols represent processes and the data flow through the system is presented. It highlights that Gen2Corr and TwoDist are the two main functions which are built within the RCM.

Gen2Corr

It is shown that the RCM requires four different types of input. I.e. Probability distributions, relevant parameter values, the defined population correlation coefficient, and the required sample size.

The last two inputs will be used to begin the data flow into Gen2Corr, and the output from Gen2Corr will be two sets of correlated random numbers from U[0,1). These two outputs are called X0(1..n) and Y0(1..n).

Gen2Corr represented the date flow and routine of the conceptual approach that was discussed in the previous chapter, Chapter 3. I.e. After the initial swapping the final adjusted arrays are X(1..n) and Y(1..n). These have perfect sample means and s.d.s.

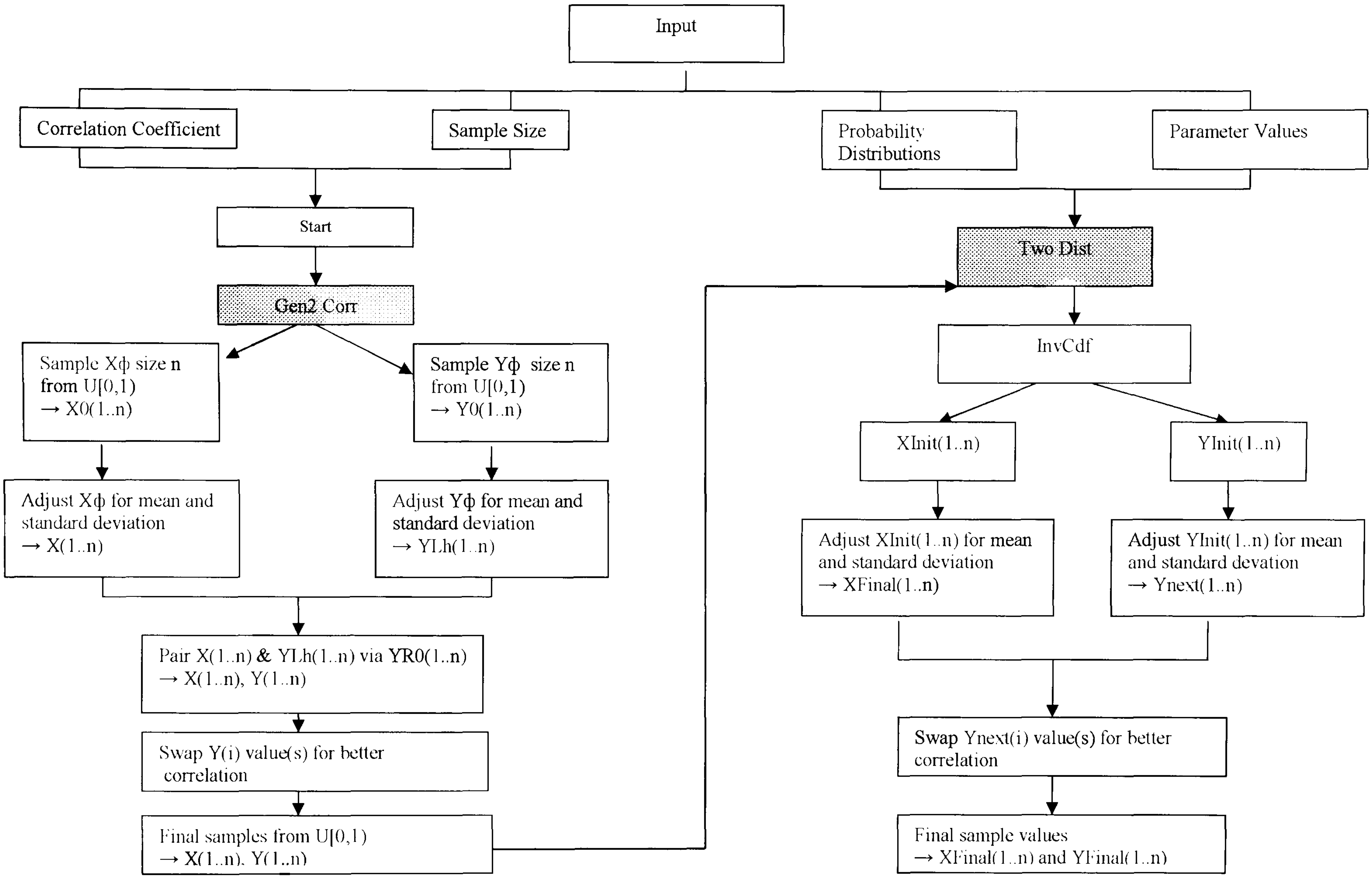


Figure 4.1 Flow chart showing the construction of the RCM

TwoDist

TwoDist is a function to generate sample values from two assigned probability distributions (in particular, their c.d.f.s) with corresponding input for parameter values. The two sets of correlated random numbers generated in Gen2Corr enter TwoDist as additional inputs to start the process.

Approach used in TwoDist:

1) Requires the two U[0,1) samples X(1..n) and Y(1..n) generated by Gen2Corr and inputs the two p.d.f.s and required sample size.

2) Transforms the assigned probability distributions into cumulative form in InvCdf

3) Uses the output from Gen2Corr in conjunction with the two c.d.f.s to identify the initial sample values from the input distributions with – at this stage – only an approximate value for the product moment correlation coefficient. These two arrays are currently called XInit(1..n) and YInit(1..n).

4) Adjusts the initial sample means and standard deviations to match with the expected values corresponding to the assigned probability distributions. Now these values are stored in XFinal(1..n) and YNext(1..n).

5) Swaps the adjusted sample values if necessary until their correlation coefficient is acceptably close to the desired correlation coefficient. Only the YNext(i) values may be altered, so that these values are finally stored in YFinal(1..n).

At the end of the process in TwoDist, two sets of sample values will have been generated. They are not only correlated, but the shapes of the frequency polygons or histograms constructed using these sample values will be very similar to the assigned probability distributions. This will be demonstrated using descriptive statistics generated in the RCM.

The inclusion of expanded probability distributions into TwoDist will be discussed in Chapter 5.

4.3.2. Sequence diagram

This section produces a Sequence Diagram, Figure 4.2, where the data flow of this RCM is shown. The purpose of creating a Sequence Diagram is different from that of creating a flow chart. A flow chart is where the steps that the programme will be going through in the RCM are presented, whereas the sequence diagram is showing:

1) the through flow of the data;

2) the place where the data are stored after each process;

3) the output from each step;

4) where the input is coming from; and

5) where the output is going to

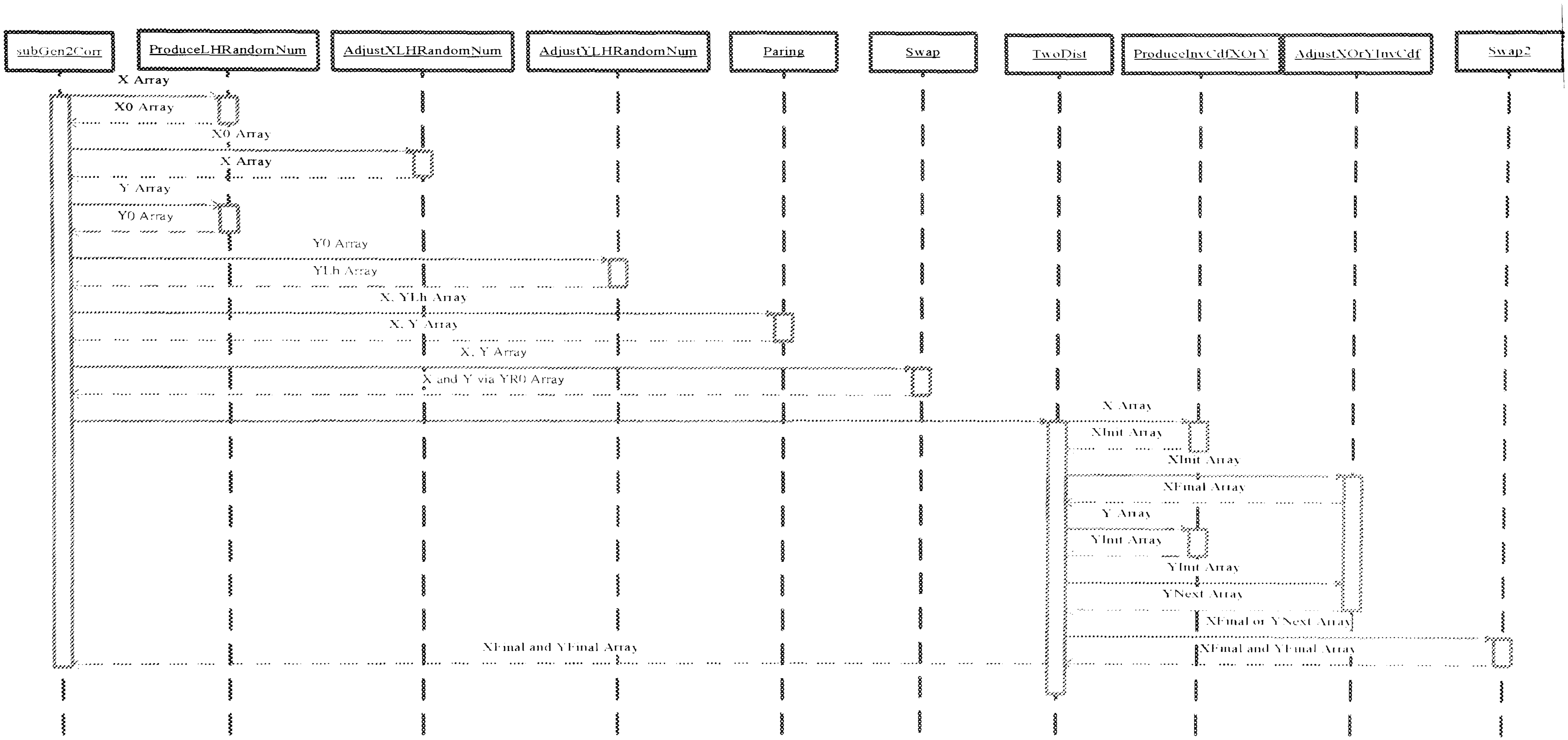


Figure 4.2 Sequence diagram showing data flow in the processes of the RCM

subGen2Corr

subGen2Corr generates two sets (to be stored in the X and Y arrays) of correlated random numbers from U[0,1). The four individual processes involved here are ProduceLHRandomNum to Swapping below.

ProduceLHRandomNum

The RCM does not use pure random numbers from a built-in random number generator. In fact, these random numbers are generated in a function called ProduceLHRandomNum where a sample of n random numbers is generated using the Latin hypercube sampling method and returns them to Gen2Corr.

AdjustXLHRandomNum

The final array generated by Gen2Corr will be passed to AdjustXLHRandomNum for adjustment so that the expected mean and variance from these n sample values are transformed to 1/2 and 1/12 respectively, since these sample values should reflect a Uniform distribution. These adjusted values will then be returned to Gen2Corr as the array X(1..n).

When it is done, the same process will be gone through for random numbers which will be stored in YLh(1..n).

Pairing

Gen2Corr now contains the X and YLh arrays which contain the adjusted random numbers of X and Y from 0 to 1 respectively. These values are passed to the function named Pairing where the shuffling, sorting and ranking of array YLh via YR0(1..n) takes place for the purpose of correlating array YLh with array X to get a good degree of relationship between them. The resequenced YLh array is now called Y(1..n).

Swapping

Now the n values in the arrays X and Y are ready for the function in Gen2Corr called Swapping where the sequence in array Y will be rearranged but the array X will remain unchanged so that the product moment correlation between these two sets of data are as close to the required correlation as possible. They are updated in arrays X(1..n) and Y(1..n) by Gen2Corr for the next use.

<u>subTwoDist</u>

As mentioned before, the final output from Gen2Corr will be used to join the initial input in TwoDist. The output from Gen2Corr was the two sets of random numbers which have the Uniform distribution's property and have been adjusted for mean, standard deviation and correlation coefficient with an acceptable level of precision and an approximate correlation. These arrays are X(1..n) and Y(1..n). The four processes comprising subTwoDist are described below.

ProduceInvCdfX

The X(1..n) values are passed to the function called ProduceInvCdfX which involves transforming the assigned probability distributions into cumulative form. This situation is most relevant when, of course, the distributions of the two variables are more complex than U[0,1), as described in Chapter 5. The transformation process will then map the array X against the corresponding c.d.f. to get the values. At the end of the process, these sample values will be stored in TempInvCdfArray and are available as output from TwoDist.

ProduceInvCdfY

Here the values in the array Y are treated in the same way that the values of X were in ProduceInvCdfX.

AdjustXInvCdf

To ensure the sample values in XInit(1..n) are truly representing the assigned probability distribution, they need to be adjusted for mean and standard deviation. This is done in the function named AdjustXInvCdf. The resulting array is called XFinal(1..n).

AdjustYInvCdf

Similarly, to ensure the sample values in YInit(1..n) are truly representing the assigned probability distribution, they need to be adjusted for mean and standard deviation. This is done in the function named AdjustYInvCdf. The resulting array is called YNext(1..n).

Swap2

Finally, the adjusted values of YNext are swapped until the correlation coefficient between these two sets of data is acceptably close here, and these values are stored in the array YFinal. The final outputs from Swap 2 are returned to TwoDist and are then ready to be used. The arrays are XFinal and YFinal.

A summary of the functions used in the RCM, together with the objectives, input and output from each function, are tabulated in Table 4.2 below.

Function Name	Task	Input	Output
subGen2Corr	Store returning values which are between 0 ~ 1 will be used as probability in TwoDist	Not applicable	Not applicable
ProduceLHRandomNum	Generate random numbers using Latin Hypercube method	Sample X(1..n) or Y(1..n) from U[0,1)	X0(1..n) or Y0(1..n)
AdjustXLHRandomNum AdjustYLHRandomNum	Adjust X or Y array for mean and standard deviation so that they have perfect means and s.d.s.	X0(1..n) or Y0(1..n)	X(1..n) or YLh(1..n)
Pairing	Sort and rank array YLh so that X and YLh have some sort of relationship	X(1..n), YLh(1..n)	X(1..n), Y(1..n)
Swap	Rearrange the order of Y via YR0(1..n) so that their relationship is closest to required	X(1..n), Y(1..n)	X(1..n) and rearranged Y(1..n)
subTwoDist	Store returning values which have been assigned for probability distributions	Not applicable	Not applicable
ProduceInvCdfX ProduceInvCdfY	Use the X(1..n) or Y(1..n) output from Gen2Corr to generate a sample with the required probability distribution	X(1..n) or Y(1..n), the given pdfs, relevant parameters values	XInit(1..n) or YInit(1..n)
AdjustXInvCdf AdjustYInvCdf	Adjust XInit(1..n) or YInit(1..n) for mean and standard deviation	XInit(1..n) or YInit(1..n)	XFinal or YNext
Swap2	Rearrange the order of terms within YNext so that the correlation with X(1..n) is acceptably close to the required correlation	XFinal(1..n), YNext(1..n)	XFinal(1..n) and YFinal(1..n)

Table 4.3 Tabulated functions in the RCM

Summary

This chapter has further demonstrated the conceptual approach discussed in Chapter 3 with a full illustrative step by step example. As the conceptual approach needs to be turned into a computer model for the purpose of testing the algorithm formulation designed in this research, the second part of this chapter presented the conceptual approach in flow chart and sequence diagrams. For each process is explained its function, the input required by the process, and the output arising from the process.

There are two components in the computer model i.e. Gen2Corr and TwoDist. The conceptual approach discussed in Chapter 3 and first part of Chapter 4 is encapsulated within Gen2Corr. The main difference between Gen2Corr and TwoDist is that the latter not only targets U[0,1) distributions in the formulation, but also other probability distributions such as Triangular, Normal and Beta distributions.

Consequently, Chapter 5 will explain specifically the ProduceInvCdf process which includes how the non-analytical inverse density function is built into the RCM, and how the swapping process is incorporated.

Chapter 5: Implementing the Correlation Model to Include Other Probability Distributions

5.1 Introduction

In Chapter 4, an approach was demonstrated for generating a set of correlated pairs of random numbers, each variable having the underlying Uniform distribution U[0,1). An algorithm required for achieving this objective is written in the Gen2Corr function in the computer model.

The algorithm has been extended to a more general probability distributions, i.e. the general Uniform, Triangular, Normal and Beta distributions. The process of transforming the two sets of random numbers onto these distributions is written in the TwoDist function in the computer model.

The process built in TwoDist is presented in Figure 5.1 below. The mathematics of the transformation and how it is handled in programming will become the main discussion issues in this chapter.

The input into TwoDist consists of (1) n pairs of numbers between 0 and 1 from U[0,1) called X(1..n) and Y(1..n), such that their sample product-moment correlation coefficient has been adjusted by Gen2Corr to be acceptably close to the required correlation coefficient, together with (2) the types and parameters of the key distributions to be modelled. The InvCdf function matches the values of X(1..n) and Y(1..n) against the respective c.d.f.s, which either are analytical functions or have to be generated using numerical integration (see below).

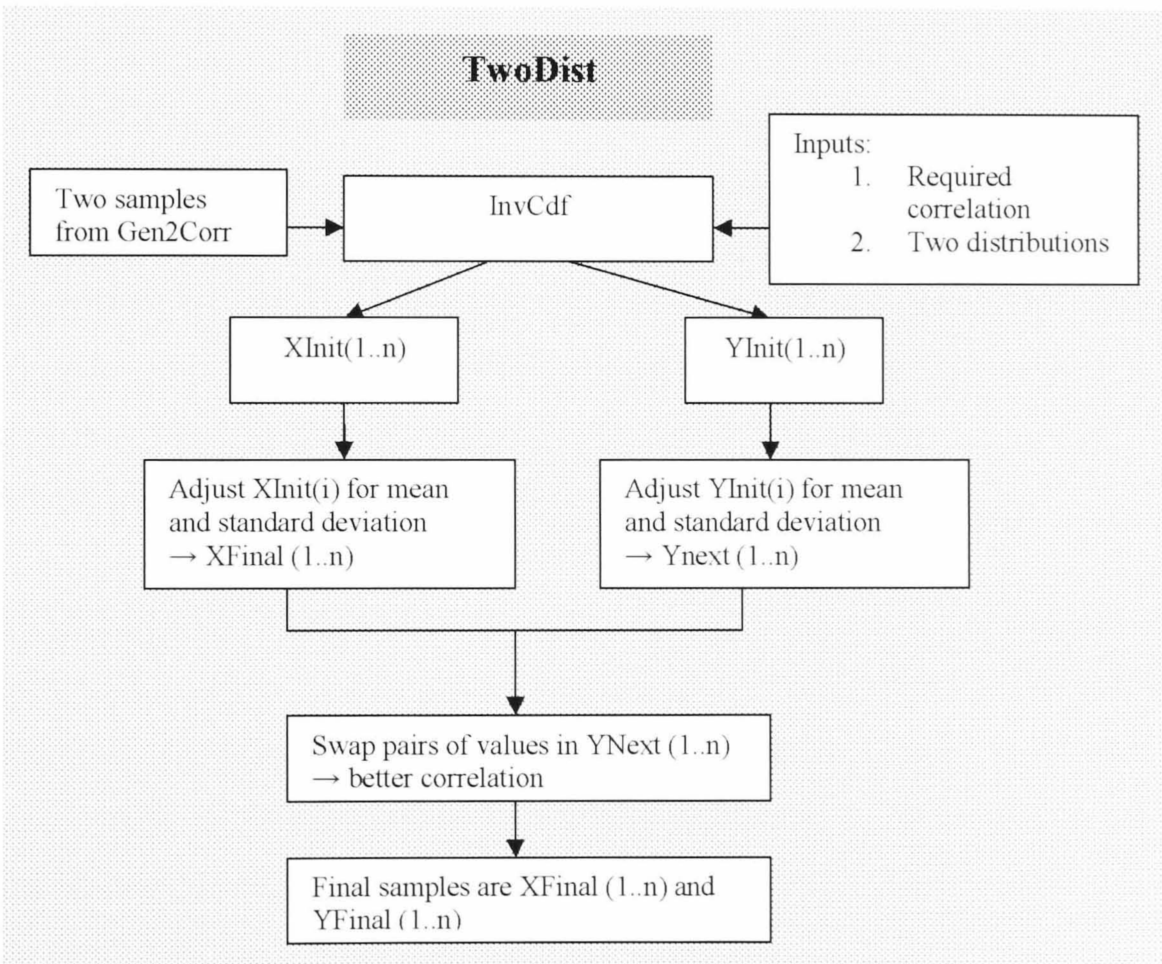


Figure 5.1 TwoDist flow chart

At the end of the InvCdf routine two samples, XInit(1..n) and YInit(1..n), either having any of the selection of probability distributions available in the RCM, are generated. At this stage their sample means and s.d.s are not necessarily equal to their population parameter values, and their sample product-moment correlation coefficient is only a first approximation to the target value.

These two sets of sample values are then adjusted for mean and s.d., resulting in arrays XFinal(1..n) and YNext(1..n), so that they accurately represent the required probability distributions. Note that this does not guarantee that other key parameters are accurately represented, e.g. skewness, kurtosis and higher moments, but in practice the skewness and kurtosis are reasonably acceptable values, and in general their accuracy appears to increase with sample size.

The last step in the TwoDist function is swapping selected pairs of values from YNext(1..n) so that the correlation coefficient of the final two arrays is as close as possible to the required product-moment correlation coefficient. This is discussed in section 5.3 below.

These final two samples are called XFinal(1..n) and YFinal(1..n). At this stage for both of the samples, their sample mean and standard deviation will be equal to the expected values of the parameter of the input distribution, and their product-moment correlation coefficient will be a very good approximation to the required product-moment correlation coefficient.

5.2 Generating sample values using the InvCdf function

This section describes the algorithm in terms of the mathematics and programming that are built into the InvCdf routine. The purpose of this routine is to generate samples X and Y using the correlated random numbers generated in Gen2Corr, called X(1..n) and Y(1..n), onto the assigned probability distributions. Within InvCdf the general element (i^{th}) of these two input arrays is called Init[i].

Table 5.1 below shows how the inverse density function built in the TwoDist function transforms Init[i] into each of two sets of sample values with which the distribution is assigned. Pseudo code has been used to demonstrate the various functions. The notations used in this table are explained below.

Probability Distribution	Formulae for Simulating Values
Uniform (min, max)	Min + (max – min) * Init[i]
Triangular (min, mode, max)	if Init[i] < (most likely – min) / (max – min) Min + {Init(i) * (max - min) *(mode - min)}$^{1\ 2}$ Else Max – {(1 – Init[i]) * (max – min) * (max – mode)}$^{1\ 2}$
Normal (mean, std_deviation)	For j = 1 to n If Init[i] between CDFN[j] and CDFN[j+1] a + (b – a) * zx end if next j where a and b are the class boundaries of the j^{th} class zx = fraction between j and j+1
Beta (least, mode, greatest)	For j = 1 to n If Init[i] between CDFB[j] and CDFB[j+1] a + (b – a) *zx end if next j where a and b are the class boundaries of the j^{th} class zx = fraction between j and j+1

Table 5.1 Formulae for simulating values from the four distributions

In the above table, the inverse density functions for Uniform and Triangular distributions are standard book formulae. It is a clear cut mathematical calculation to derive the cumulative curve for the Uniform distribution i.e. simply fit the generated Init[i] into the standard formulae.

The only additional step for generating the cumulative curve for the Triangular distribution is that prior to those steps as for Uniform distribution, it is necessary to identify if Init[i] lies within the left or right part of the p.d.f so that the appropriate formula is used. Because of the straight forward process in both the Uniform and Triangular cases, this will not be discussed further in this chapter.

However, the Normal and Beta distributions require specially designed routines because their c.d.f.s are not analytic functions.

To construct a cumulative curve for both Normal and Beta distributions, calculate and accumulate the area under the curve when it is divided into n strips. This follows the well-known Simpson's rule, which is a method of approximate numerical integration, equivalent to assuming that the curve being integrated is the same as a series of piecemeal quadratic curves with the same endpoints and midpoint for the n-strips into which the area is divided. The flow of constructing a cumulative curve for Normal and Beta distribution using Simpson's rule is presented in Figure 5.2 below. Details in each step will be discussed below.

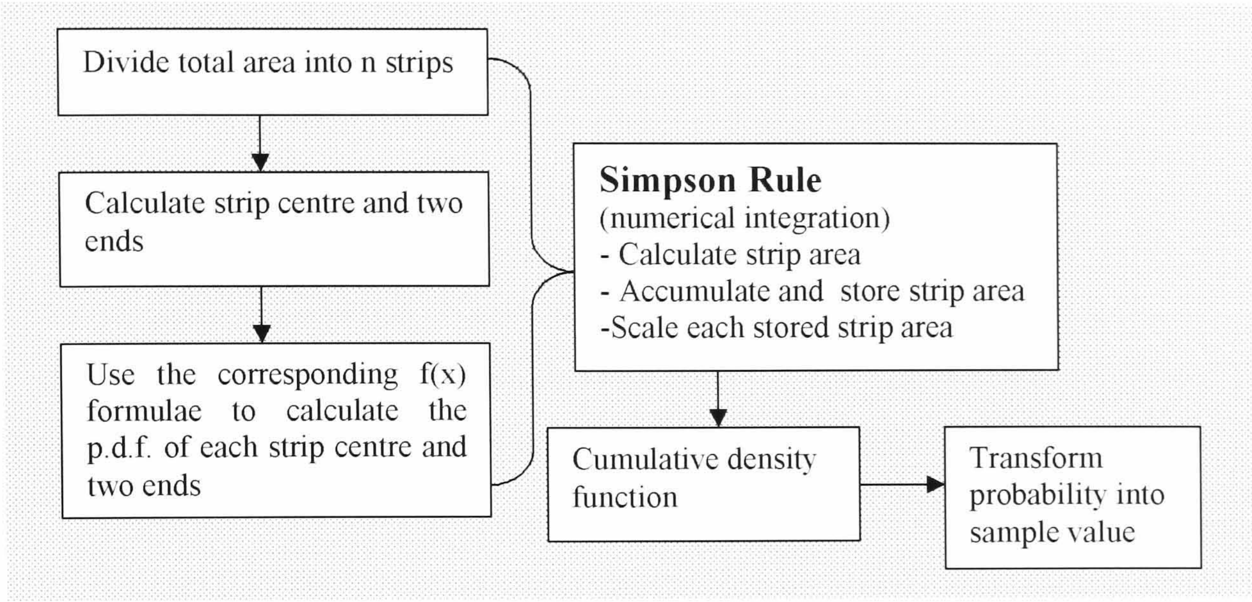


Figure 5.2 Simpson's Rule process

5.2.1. Modelling the Normal distribution in subGetNormalDistribution

The purpose of subGetNormalDistribution is to generate the c.d.f. F(z) if z is N(0,1), for z from -4.0 to 4.0 in 1000 equal intervals. Hence cdf(0), cdf(500), and cdf(1000) are equal to 0.0, 0.5 and 1.0 respectively.

For example, cdf(745) = 0.975 because z = -4.0 + 745 * 8/1000 = 1.96, and $\Phi(1.96) = 0.975$. The choice of 1000 as the number of intervals was made after some experimentation, and the calculated values were tested to be acceptably close to values tabulated elsewhere.

To form the c.d.f. of a Normal distribution, the area under the curve will be calculated and accumulated. Initially, a standard Normal distribution is used, for which the total area on the right and left are symmetric. Therefore, this function will only calculate the c.d.f. for the positive half i.e. z > 0.

Simpson's rule is applied here to integrate the area under the curve into 1000 strips so that, if a typical strip lies between x = a and x = b:

$f(x)dx \approx (b-a)\ [\ f(a) + 4f(\tfrac{1}{2}\{a+b\}) + f(b)\ ]\ /\ 6$

Because great accuracy is required, the strip limits are calculated as double precision numbers, for which the strip centre is vi#, and the two ends are viLeft# and viRight#, and their p.d.f. values are fi#, fileft# and firight# respectively calculated using:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

```
w#  = 0.004
hw# = 0.002
k0 = 1 / Sqr(2 * pi)

For i = 1 To 1000
  Vi# = i * w# - hw#: viLeft# = vi# - hw#: viRight# = vi# + hw#
  fi# = Exp(-0.5 * (vi^2))
  fileft# = Exp(-0.5*(vileft#)^2)
  firight# = Exp(-0.5*(viright#)^2)

   StripArea# = k0 * w * (fiLeft# + 4 * fi + fiRight#) / 6

   TempInvCdf(i) = TempInvCdf(i - 1) + StripArea#

Next i
```

The frequency of the right hand side of the distribution from z = 0 to 4 and the range is divided into 1000 strips to become i = 1 to 1000. Each strip is therefore of width w# = 0.004, and hw# = 0.002 is half of the strip width. The choice of "±4 s.d.s" was made because the probability of getting a value which is more than 4 s.d.s from the mean is negligible, whereas the probability of getting a value which is 3 s.d.s from the mean or more is not neligible.

There are two constants in the formula for the p.d.f. f(x) above:

The first constant $\frac{1}{\sqrt{2\pi}}$, k0, ensures that the total area under the p.d.f. curve is 1.0; and the second constant = e = Exp (1).

Standardising an x value yields:

$z = (x-\mu)/\sigma$, which is also the value of vi# or vileft# or viright#, as appropriate.

To account for $-\frac{1}{2}[(x-\mu)/\sigma]^2$, calculate fi#, fileft# and firight# which are the p.d.f. values of the strip centre and two ends respectively.

Now calculate the strip area using Simpson's Rule:

```
StripArea# = k0 * w * (fiLeft# + 4 * fi# + fiRight#) / 6
```

Accumulate each strip area with the previous accumulated strip areas and store in TempInvCdf(i).

```
TempInvCdf(i) = TempInvCdf(i - 1) + StripArea#
```

To make sure the final c.d.f. i.e. TempInvCdf(1000) value = 1.0 exactly, scale each area stored in TempInvcdf(i). Assume the final c.d.f. is t:

```
t = TempInvCdf(1000)
For i = 1 To 1000: TempInvCdf(i) = TempInvCdf(i) / t:
Next i
```

(Note that t should be very close in value to 1)

To find the corresponding c.d.f. for Z between -4 to 4, the above 1000 values of TempInvCdf(i) are now to include the left hand side of the distribution i.e. $z < 0$. As a result, the area of each strip is twice the initial area, so that the area stored in TempInvCdf(i) becomes the area stored in TempInvCdf(2 * i).

Thus, for values of $z > 0$:

```
CDFN(0) = 0: CDFN(500) = 0.5: CDFN(1000) = 1
For i = 1 To 500: CDFN(500 + i) = TempInvCdf(2 * i): Next i
```

Whereas for values of $z < 0$, symmetry yields:

```
For i = 1 To 500: CDFN(i) = 1 - TempInvCdf(1000 - 2 * i): Next i
```

At the end of this routine, the cumulative distribution of a standardised Normal distribution will have been constructed.

Finally, use the constructed curve to find a sample value from a Normal distribution, where Init[i] is the probability.

Scan the range to find the strip that the assigned cumulative probability falls into, then calculate the corresponding sample value where:

a = the lower class boundary of the j^{th} class,
b = the upper class boundary of the j^{th} class,
zx = fraction between j and j+1 to get the exact sample value in the range

The pseudo code for this is:

```
For j = 1 to n
  If Init[i] between CDFN[j] and CDFN[j+1]
   a + (b – a) * zx
  end if
next j
```

5.2.2. Modelling Beta distributions

There are various ways of defining the parameters of a Beta distribution. Of course the smallest and largest possible values must be specified: 'a' and 'b'. Other possible parameters are the modal value, the expected (i.e. mean) value, the median (Q2), and either or both of the two shape parameters v and w.

In general, if the variable in question, x, lies in $a \le x \le b$ and the Beta distribution is then defined in terms of the shape parameters v and w, the p.d.f. of x is
$f(x) = [\ (x-a)^{v-1}(b-x)^{w-1}\ ] \div [\ (b-a)^{v+w-1}\beta(v,w)\ ]$, where $\beta(v,w) = \Gamma(v)\Gamma(w) \div \Gamma(v+w)$,

and the Gamma function, $\Gamma(\theta)$, is defined to be equal to $\int_0^{\infty} e^{-x}x^{\theta-1}dx$.

It can be shown by differentiation that the modal value of this distribution is $m = a + (b-a)(v-1)/(v+w-2)$, and the mean value is $\mu = a + (b-a)v/(v+w)$.
(See section 5.2.2a. below).

Hence standardise the value of x: $x \rightarrow (x-a) / (b-a)$, so that $0 \le x \le 1$.

This standardised Beta function is not unique (unlike the standardised Normal function), because it will vary with the shape (i.e. skewness and kurtosis) of the distribution.

In practice it does not seem sensible to ask a user to recognise from the shape of the required Beta distribution what the values of v and w are to be. It is, however, reasonable to state where the modal value, m, is.

Thus, the approach here is to specify the values of a, b and m only, not even requiring the value of μ. Then by judicious choice of either v or w, it is possible to calculate the other shape parameter and the value of μ.

A desirable property is that the slope of the p.d.f. function, f(x), tends to zero as $x \to a$ or $x \to b$ in the general Beta distribution, or as $x \to 0$ or 1 in the corresponding standardised distribution. It will also be shown that in this case both v and w should be > 2. (See section 5.2.2a. below).

Suppose now that the modal value of x in $a \le x \le b$ is closer to a than to b, so that $m < (a+b)/2$, and thus $a + (b-a)(v-1)/(v+w-2) < (a+b) / 2$.

$$2a(v+w-2) + 2(b-a)(v-1) < (a+b)(v+w-2)$$
$$2(b-a)(v-1) < (v+w-2)(a+b-2a)$$
$$2(b-a)(v-1) < (v+w-2)(b-a)$$

Now, $a < b$, so that $b - a > 0$. Divide this expression throughout by (b-a) yielding $2(v-1) < v + w - 2$, which simplifies to $v < w$.

$\therefore$ If the modal value, m, is closer to the left hand limit a than to b, it has been shown that $v < w$, and that there is positive skewness.

If $m < \mu$ then $a + (b-a)(v-1)/(v+w-2), < a + (b-a)v/(v+w)$, where $b - a > 0$, so that $(v - 1)(v + w) < v(v + w - 2) \to -(v + w) < -2v$, or $v < w$.

Hence for positive skewness: $m < \mu < (a+b) / 2$, and $v < w$, where v and $w > 2$.

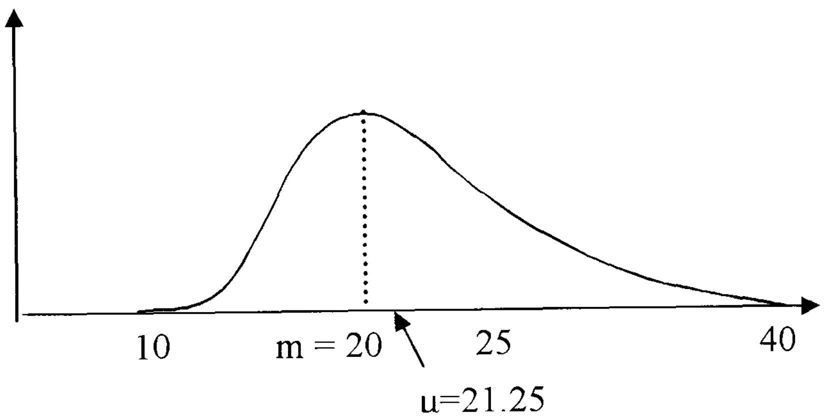
So, if the chosen Beta distribution is to have positive skewness, choose a convenient value of v which is greater than 2, such as $v = 3$, say. The value of w will be larger than v and can be deduced from the formulae for m:
$m = a + (b-a)(v-1)/(v+w-2)$.

Putting $v = 3$, $m = a + (b-a)(2)/(w+1)$. Then $w+1 = 2(b-a) / (m-a)$, so that $w = [\ 2(b-a) - (m-a)\ ] / (m-a)$, $= (2b-m-a) / (m-a)$.

For example, if $a = 10$, $b = 40$ and $m = 20$ then $w = (80-20-10)/(20-10)$, $= 5$, so that $v = 3$ and $w = 5$ (with $v < w$).

The value of μ would then be calculated to be
$\mu = a + (b-a)v/(v+w) = 10 + 30*3 / (3+5)$, $= 10 + 90/8$, $= 21.25$, and this is $> m$.

Note that the mean value is greater than the modal value, but only just.



A question might be asked: "why would one not specify the values of a, b, m and μ, so that one states μ rather than v?"

For example, $a = 10$, $b = 40$, $m = 20$ (as above) and $\mu = 25$. This makes μ rather larger than m, but it also will yield non-desirable values of v and w:

"$m = a + (b-a)(v-1)/(v+w-2)$" gives $20 = 10 + 30(v-1)/(v+w-2)$, $\rightarrow 2v - w = 1$.

Also, "$\mu = a + (b-a)v/(v+w)$" yields $v + w = 3v$, so that $w = 2v$, $\rightarrow v = 1$, < 2.

This demonstrates that, although from the shape of the Beta curve it is easy to identify the values of a, b and m, it is not an easy skill to estimate the expected value. It is easier, of course, to standardise the distribution so that $0 \le x \le 1$. This is easily achieved via the transformation $x \rightarrow (x\text{-}a) / (b\text{-}a)$.

For this particular standardised Beta distribution the modal and mean values are then m0 and μ0, say, where m0 = (v-1) / (v+w-2) and μ0 = v / (v+w).

In the above example, where a = 10, m = 20 and b = 30, we chose v = 3 and deduced that w = 5. Then m0 = 2/6 = 0.3333 to 4 d.p.s, and μ0 = 3/8, = 0.375. If v = 3 then m0 = 2 / (w+1), so that w = 2/m0 –1 , = 6 – 1, = 5, as before.

Then the rules for specifying the values of v and w are:

(1) If the standardised modal value, mϕ, < 0.5, set v = 3; calculate w from
w = –1 + 2 / m0

(2) From symmetry, if mϕ > 0.5, set w = 3, and calculate v from
v = (1 + m0) / (1 – m0)

(3) If mϕ = 0.5, so that the distribution is symmetric, set v = w = 3.

If all 3 cases the expected value, μ, is $\dfrac{v}{v+w}$, and is thus easily calculated.

[In all this description the smaller of v and w is always to be equal to 3. This could be any positive value which is greater than 2, so that "3" is arbitrary. For example, it could equally well be "2.5"].

5.2.2a. The standardised Beta distribution and its expected value

To standarise x, set a = 0 and b = 1, so that

$$f(x) = \frac{x^{v-1}(1-x)^{w-1}}{\beta(v,w)} \quad \text{and}$$

$$\frac{df}{dx} = \frac{(v-1)x^{v-2}(1-x)^{w-1} - (w-1)x^{v-1}(1-x)^{w-2}}{\beta(v,w)}, = 0 \text{ at the finite tails}$$

if $x^{v-2} = 0$ or $(1-x)^{w-2} = 0$ or $(v-1)(1-x)-(w-1)(x) = 0$.

When x = 0, $x^{v-2} = 0$ if $v > 2$; when x = 1, $(1-x)^{w-2} = 0$ if $w > 2$; and $(v-1)(1-x)-(w-1)(x) = 0$ if $x = \frac{v-1}{v+w-2}$, and this is where the mode occurs if $\frac{df}{dx} = 0$, as required.

The expected value of the standardised Beta distribution is $\mu\phi$, say,

$$= \int_0^1 \frac{x.x^{v-1}(1-x)^{w-1}}{\beta(v,w)} dx.$$

$$\int_0^1 f(x)dx = \int_0^1 \frac{x^{v-1}(1-x)^{w-1}}{\beta(v,w)} dx, = 1 \text{ so that } \int_0^1 x^{v-1}(1-x)^{w-1} dx = \beta(v,w).$$

$$\therefore \mu\phi = \int_0^1 \frac{x^{(v+1-1)}(1-x)^{w-1}}{\beta(v,w)} dx = \frac{\beta(v+1,w)}{\beta(v,w)} * \int_0^1 \frac{x^{(v+1-1)}(1-x)^{w-1}}{\beta(v+1,w)} dx$$

$$= \frac{\Gamma(v+1)\Gamma(w)/\Gamma(v+w+1)}{\Gamma(v)\Gamma(w)/\Gamma(v+w)} * 1 = v\frac{\Gamma(v)}{\Gamma(v)} * \frac{\Gamma(v+w)}{(v+w)\Gamma(v+w)} = \frac{v}{v+w}$$

$\therefore$ $\mu 0 = v / (v+w)$, as claimed earlier.

The required values to calculate the fi#, fileft# and firight# values which are the frequencies of the strip centre and two ends respectively are obtained. The calculation for f(x), i.e. $\beta(v,w) = \Gamma(v)\ \Gamma(w) \div \Gamma(v+w)$, is discussed in section 5.2.3.

To continue the process in subGetBetaDistribution, Simpson's Rule is used and each strip area is accumulated and stored in TempInvCdf(i).

```
StripArea = w * (fiLeft# + 4 * fi# + fiRight#) / (6*betavw)

TempInvCdf(i) = TempInvCdf(i - 1) + StripArea
```

To ensure the final c.d.f. value = 1.0 exactly and all the others are adjusted pro rata is achieved when generating the standardised Normal c.d.f. values, let t be the 1000^{th} c.d.f. value of the Beta Distribution.

```
t = cdfB(1000)
For i = 1 To 1000: cdfB(i) = cdfB(i) / t: Next i
```

The c.d.f. of the standardised Beta distribution has now been constructed, and to find a sample value using the constructed curve for a general Beta distribution, where Init[j] is the probability, scan the interval that the assigned cumulative probability falls into then calculate the corresponding sample value where:

a = left hand class boundary of this interval and

b = greatest,

zx = fraction between j and j+1 to get the exact sample value in the range

```
For i = 1 to n
  If Init[i] between CDFN[j] and CDFN[j+1]
  a + (b – a) * zx
  end if
 next j
```

5.2.3. Modelling a Gamma function in subGetGamma

The lower part of the formula for f(x), i.e. $\Gamma(v)\Gamma(w)/\Gamma(v+w)$, is calculated in a different function called subGetGamma.

A gamma function of a parameter x0 > 1.0 is calculated using Simpson's rule. To start the routine, reduce the parameter to between 1 and 2 if neccessary.

For example: gamma(3.7) = (3.7 – 1) * gamma (3.7 – 1)
= 2.7 * gamma(2.7)
= (2.7 * 1.7) * gamma(1.7)
= 4.59 * gamma(1.7)

This is expressed as:

```
Between1and2:
  If x0 > 2 Then x0 = x0 - 1: factr = factr * x0: GoTo Between1and2
```

Now calculate gamma(θ), for $0 \leq \theta \leq 1$, using: $\Gamma(x) = \int_0^\infty e^{-x} x^{\theta-1} dx$.

Since it is the area of a Gamma distribution that is calculated, Simpon's rule is used. To integrate a Gamma distribution from 0 to 30, say, divide the total area from x = 0 to x = 30 into 30,000 strips of width 0.001. For each strip evaluate the function in the middle, at the left and right which become vi#, vileft# and viright#. (N.B. Integrating up to $x \leq \infty$ cannot be done here, but the area to the right of x = 30 is infinitesimal, so that any value greater than 30 can be ignored).

Back to the example above, Gamma(3.7) = $4.59 * \int e^{-1.7} * x^{0.7} dx$

For simplicity, define $k = \theta - 1$. The programmed p.d.f. values vi#, vileft# and viright# are expressed as:

```
fi# = Exp(-vi) * (vi ^ k)
fiLeft#= Exp(-viLeft) * (viLeft ^ k)
fiRight# = Exp(-viRight) * (viRight ^ k)
```

At the end of the routine, the relative frequency of each strip has been calculated. Finally, each strip area is accumulated with the previous accumulated strip area.

```
For j = 1 to 30,000
StripArea# = w * (fiLeft# + 4 * fi# + fiRight#) / 6
gamma = gamma + StripArea
Next j
```

The output is the value of gamma(x). Restore to the original parameter

```
Gamma(x) = gamma * factr
```

The same process is carried out for $\Gamma(v)$, $\Gamma(w)$ and $\Gamma(v+w)$.

5.3 Improving the correlation value via swapping

This section demonstrates how the adjustment of sample Y is handled in the programming so that the correlation with sample X is made closer to what it is supposed to be. Figure 5.3 is a flow chart of the swapping process.

It can be seen that sample $\sum xy$ or named "SumXYNow" is used to compare with the required $\sum xy$ to identify the required improvement in sample Y. The difference, "ModDif", between the sample correlation coefficient and required correlation coefficient is calculated to identify if any swap is necessary in sample Y. If the answer is 'yes', then the routine will start taking two values in sample Y and swap. If the particular pair passes the conditions set in the routine, it is said that the pair is significant in the particular run and the contribution or reduction in sample $\sum xy$ will be stored, but will be overwritten subsequently if a better pair is found.

The routine will stop either when "ModDiff" is very small and insignificant, or no better pair can be swapped, or when the routine has carried out four complete runs. 'Four' is an arbitrary number but, in practice, either the process usually is completed in or before the 4th full iteration, or the improvement from the 5th full iteration normally is very small indeed. Later it is observed that a larger number of runs may sometimes be required if the sample size is large (section 6.3.4.).

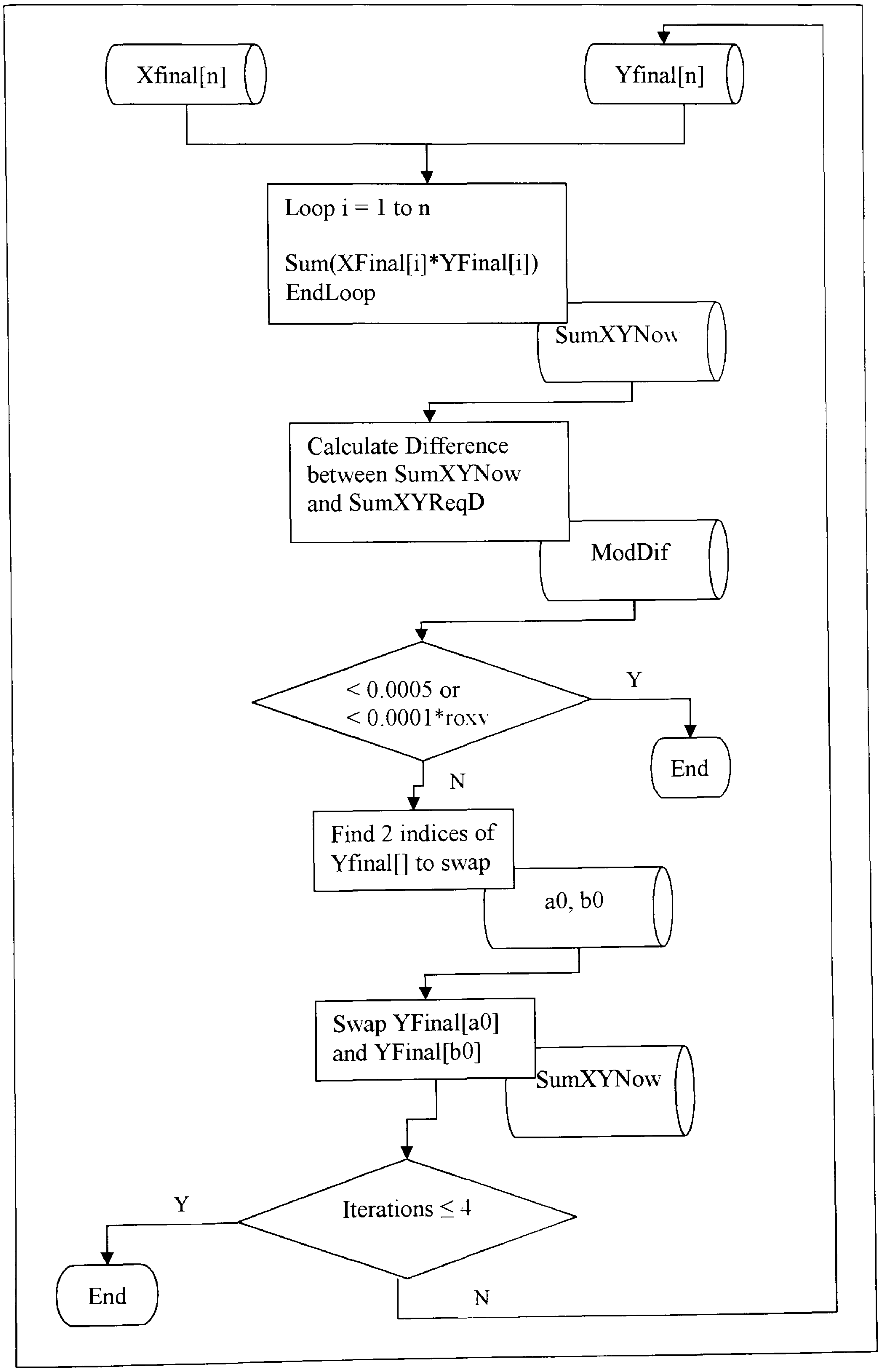


Figure 5.3 Swapping process

To start the swapping process, the sample $\sum xy$ or "sumxyNow" is calculated using the generated sample values. The required $\sum xy$ or "SumxyReqd", and is calculated as $n\rho * \sigma_x \sigma_y + n\mu_x * \mu_y$.

The sample correlation coefficient named "rxy", and is compared with the required correlation coefficient or named "roxy" in the programme.

```
For each of the paired values in turn:
    xy = XFinal(i) * YFinal(i)
    sumxyNow = sumxyNow + xy

    AvexyReqd = roxy * sigmax * sigmay + mux * muy
    SumxyReqd = AvexyReqd * n

    rxy = (sumxyNow / n - mux * muy) / (sigmax * sigmay)
    AvexyNow = sumxyNow / n
    ModDif = Abs((AvexyNow - mux * muy) / (sigmax * sigmay) - roxy)
```

The difference between the sample value of $\sum xy$ and the required value of $\sum xy$ is called "sumxychange" and its absolute value is called "Diffxy".

```
SumxyChange = (SumxyReqd - sumxyNow)
AvexyChange = SumxyChange / n
Diffxy = Abs(SumxyChange)
```

"ModDif" is the average absolute change that is sought in the product of each sampled pair.

If the "rxy" value is equal to "roxy" to four decimal places (i.e. if "ModDif" is less than 0.00005 or "ModDif" is less than 0.0001 * required correlation coefficient in absolute terms), the swapping process will cease and set the current sample values to be the final samples X and Y.

```
If ModDif < 0.00005 Or ModDif < 0.0001 * Abs(roxy)
Then GoTo GetResult
```

If the above condition doesn't hold, then this potential improvement process is repeated until it is satisfied. Initially the two values which are to be swapped are a0 and b0. Hence the process continues until either the 'improvement indicator', "intImprove", does not change from zero, or until four (arbitrary) full iterations have been completed.

```
    a0 = -1: b0 = -1: intImprove = 0: VarImprove = 0

For a = 1 To n - 1
    For b = a + 1 To n
        xa = XFinal(a): xb = XFinal(b): ya = YFinal(a): yb = YFinal(b)
        z = (xb - xa) * (ya - yb): upz = updn * z
        If upz > 0 And upz < 2 * Diffxy - 0.0001
          And Abs(upz - Diffxy) <  Abs(varImprove - Diffxy)
    Then
          varImprove = upz: intImprove = 1: a0 = a: b0 = b
        End If
    Next b
Next a
```

The routine swaps two values in sample Y in each swap. There will be $\frac{1}{2}n(n-1)$ combinations in each individual run (i.e. iteration).

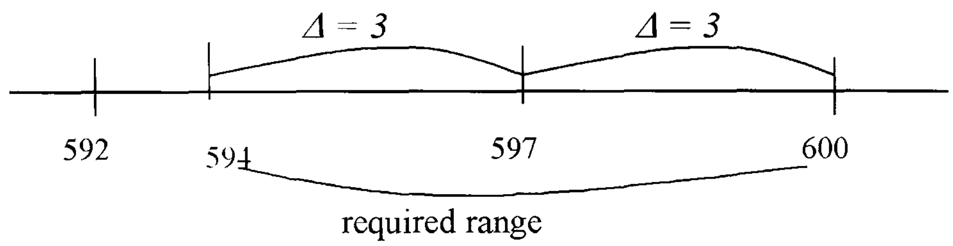
To make the z value become absolute, it is multiplied by an index called "updn" and becomes "upz".

```
If sumxyNow < SumxyReqd Then
        sUpdn = "increase": updn = 1
Else
        sUpdn = "decrease": updn = -1
End If
```

To check if the particular swap will help to improve the sample correlation, the following two conditions must be satisfied, otherwise the next possible swap will be considered:

1. There must be an improvement; and
2. The contribution or reduction must be less than twice what it is required in "Diffxy" $=\Delta$, otherwise the new approximation will be worse than the existing one.

For example, if $\Sigma xy = 600$ but the ideal required value is 597, we will wish to reduce Σxy by 3, so that $\Delta = 3$. If, then, the actual reduction is $\delta_{ij} = 8$, the revised value of Σxy would be 592, and this will not pass the test in condition 2) because the new value of Σxy would be further from the target than before. Hence this undesirable pairing will not be assigned to be the incumbent pairing.



The remaining required improvement = Abs(upz – Diffxy) must be less than the previous improvement = (varImprov – Diffxy), so that this new approximation is better than the previous one.

The initial "varImprov" is set to be varImprov = 100000 * sigmax * sigmay, so that it is impossible for the first iteration to fall into the third condition insensibly.

At the end of each complete iteration, the best pair of values of Y will be swapped and will be used for the next iteration. This is therefore a "greedy" algorithm (or "steepest descent/ascent" algorithm), and it may well be that an alternative search algorithm could achieve even better results.

After the maximum of four iterations, the programme will stop and identify the two samples in the last run as the required final paired samples X and Y. (As stated earlier, only rarely is 4 iterations not enough, and usually this occurs only if the sample size is large, such as 500).

Summary

The previous chapter, Chapter 4, demonstrated the algorithm used to generate correlated pairs of random numbers, each variable having the underlying distribution U[0,1). The generated product-moment correlation coefficient of the two samples was shown to lie within a satisfactory range.

Chapter 5 has demonstrated how this algorithm within the RCM is actually built into processes and routines within Gen2Corr.

The computer coding description has also been extended from the base approach involving two U[0,1) distributions to dealing within the routine TwoDist with a selection of other more general distributions: the general Uniform, Triangular, Normal and Beta distributions.

The general Uniform distribution was chosen because it's the simplest generalisation and has a simple analytic inverse cdf function. The general Triangular distribution also has an analytic inverse function, albeit in two sections on either side of the mode. Unlike the general Uniform distribution, however, the transformation from the U[0,1) sample generated in Gen2Corr does not have its sample mean equal to the expected value, so that both the sample mean and s.d. will need to be adjusted. The general Normal distribution was chosen because it does not have an analytic cdf and therefore its inverse cdf does not exist as an analytic function.

However a convenient property is that all general Normal distributions can be standardised to a unique N(0,1) distribution, so that the inverse cdf values calculated within the RCM could be easily benchmarked against readily available tables. The general Beta distribution was chosen because it is under-used in practice (i.e. it probably should be used more often because of its great flexibility) and, although its cdf is not an analytic function (and thus its inverse is also not analytic), the restriction in every case of either $v = 3$ or $w = 3$ ensures that inverse mappings are readily easy to calculate, and then can be benchmarked against tables of incomplete beta functions.

The results arising from a large number of balanced trials using the RCM are shown in the next chapter. In each trial the two variables are each assigned one of these four general probability distributions, for a variety of parameter values.

Chapter 6: Testing and Verifying the Validity of the RCM

6.1 Introduction

The primary aim of this chapter is to demonstrate and illustrate the level of accuracy of the output from the RCM. Thus, section 6.2 below examines the four general distributions incorporated into the RCM, verifies the accuracy of the way in which they have been modelled, and assesses the accuracy of the sampled output.

The results of key tests are presented in this section 6.2, and the output from computer runs of the RCM which underpin this section form Appendices I to VII.

Section 6.3 then – more significantly – assesses the precision of the final correlation coefficients between the samples, and the output from a range of experiments shows clearly that real and significant progress is indeed achieved. Some representative results are shown in section 6.3.2 and section 6.3.3, and the tabulated results from further combinations comprise Appendix VIII.

Finally, in section 6.4 attention turns to comparing the results of the RCM with those obtained by carrying out corresponding runs of @RISK and, in the area of modelling, it is demonstrated clearly that the RCM returns greatly improved precision. Again representative comparisons are presented here (in section 6.4) in tabulated form, and Appendix IX contains further comparisons to support the conclusions drawn here.

6.2 Checking the modelling of, and output from, the four general distributions

6.2.1. Introduction and methodology

The RCM enables variables to be represented which have any of four separate probability distributions of occurrence: the general Uniform, Triangular, Normal and Beta distributions. In each case only variables with continuous p.d.f.s have been included. Variables in the first two cases above have inverse c.d.f.s which are analytic functions. In both the Normal and Beta cases this is not true, so that solving F(x) = k for a given value of k requires tables of values of the c.d.f., F(x), to be compiled within the programme as discussed in Chapter 5.

Other popular distributions could have been included as p.d.f. types, such as the Lognormal and exponential distributions, but it was not an objective to provide here a RCM which can cope immediately with any category of distribution. Rather, the focus was to be to demonstrate the achievement for a typical selection of p.d.f. types, so that the correlation algorithm can be incorporated into a programme and use it with other p.d.f. types if desired.

It should be noted that, although the sampling of the arrays from the two distributions is carried out by the RCM using a particular form of Latin hypercubes, the correlation algorithm is equally applicable when the sampling is carried out by other means, for example using Monte Carlo Simulation.

The purpose of this section is to examine each of these four p.d.f. types and to demonstrate that in each case the sample of simulated values has a prefect sample mean and standard deviation and that the "shape" of the sample is acceptable. Two possible alternative ways of assessing the shape are (i) to compare the skewness and kurtosis of the simulated values with the corresponding expected parameter values; or (ii) to carry out a goodness of fit test. Effectively the second alternative, in the form of a χ^2 test (Hoel, 1984), is at least as powerful as the first, since a non-significant result would imply that both the measured lack of symmetry and the degree of peaked-ness are acceptable. In addition the χ^2 tests are more popular in practice than the test of skewness and kurtosis.

Thus for each of the four p.d.f. types, a set of parameter values was chosen, and then 100 values of the variable were generated. The parameter values chosen were U[18,38), T(10,30,40), N(100,62), and B(12,15,30) respectively. The analysis in sections 6.2.2. to 6.2.5. will conclude that the modelling of these four distributions is highly acceptable. Consequently little would be gained either by carrying out more than 25 runs or generating larger samples, such as 500. In the case of samples of size 10, the sample means and standard deviations were analysed and were found to be perfect, but a χ^2 test would be inappropriate here, because the total frequency is only 10. A Kolmogorov-Smirnoff test could be carried out (Hoel, 1984), but would appear to be of only minimal value. The One-Sample K-S test is a non-parametric test which compares the expected cdf of the variable with its observed cumulative frequencies.

Like other non-parametric tests, the K-S test is only slightly less powerful with large samples than parametric tests such as χ^2 tests, but the K-S test can be used with small samples – albeit with caution – whereas the χ^2 test cannot be used at all with small samples. The parameter values chosen here are fairly random and, in the case of the Triangular and Beta distributions, illustrate negative and positive skewness respectively.

In each case a total of 25 such runs were carried out. The 100 generated values per each of the 25 runs are included in the four tables comprising Appendix I to IV, the values being sorted for convenience into ascending order.

The analysis in each of the following four subsections shows for the first run only of the 25 runs how the sample mean and s.d. are calculated, and compares them with the theoretical expected value and its s.d. The 100 observed values simulated in the run are then investigated to see how many fit into each of 10 equi-probable (and exhaustive) sub-ranges of the overall range of the values the variable could take.

For example, in section 6.2.2. the variable has a U[18,38) p.d.f. Therefore splitting this range into 10 equi-probable sub-ranges will produce intervals [18.0000,20.0000), [20.0000,22.0000), . . . , and [36.0000,38.0000), with all recording of results and analysis being done to 4 d.p. accuracy, so that the expected frequencies in each of these 10 intervals will be 100 ÷ 10, = 10.

One final point needs to be made before progressing to the analysis of these four cases. When the initial sample of 100 values is generated in the TwoDist routine, the use of Latin hypercubes means that one value is generated in each of the 100 equi-probable intervals. Hence at this stage a χ^2 test should give a perfect fit, with the χ^2 value being zero.

However, the linear transformation required to adjust the sample mean and s.d. to become equal to the corresponding parameter values implies that each simulated value may be shifted – either increased or decreased – by a small amount. Consequently the observed frequencies within the ten classes could vary slightly from 10. i.e. Each one could become either 9 or 11, and it is possible to visualise the case where an observed frequency could increase to 12 or even decrease to 8.

6.2.2. Results from the general Uniform distribution U[a,b)

In this section the Uniform distribution used is U[18,38), so that the least possible and greatest possible values are 18 and 38 respectively, and the variable is equally likely to take any value between these two limits.

Note again that the consequence of "[" and ")" is of extreme mathematical interest and relevance only, so that theoretically the least value, 18 exactly, could be generated as a simulated value, but the largest possible simulated value cannot quite attain 38, but will be as close to – but less than – the exact value 38 as the particular computer being used will allow. To all intents and purposes, therefore, there is no practical distinction between U[18,38), U[18,38), and U[18,38].

The expected value is $\mu = (a + b) \div 2, = 56 \div 2, = 28$ exactly,
and the s.d. is $\sigma = (b-a) \div (12\ )^{1\ 2}, = 20 \div (12\ )^{1\ 2}, = 5.7735$ (to 4 d.p.s).

The 100 values per run are displayed in the first 100 cells in the respective column of the 25 columns (i.e. runs) in Table 1 in Appendix I. Thus the 100 simulated values in the first run are 18.1588, 18.2414, . . . , 37.8795, and form the contents of cells A1 to A100 in this table.

At the base of this column cell A104 contains the sample mean which is calculated via the Excel function AVERAGE(A1..A100), and is 28.0000. Similarly the s.d. of the 100 values is calculated via STDEVPA(A1..A100), and is 5.7735.

Hence the sample mean and standard deviation are perfect.

The c.d.f. for this distribution for values of x from 18 to 38 is $F(x) = (x-18) \div 20$.

Hence the upper class boundaries of the 10 equi-probable sub-intervals of [18,38) can be found by solving $F(x) = d \div 10$, for $d = 1, 2, \ldots, 10$, and so they will be 20.0000, 22.0000, and so on, to 38.0000.

Table 6.1 following shows for each class its lower and upper class boundaries, the first and last members observed from the table to lie in this class, and therefore the observed frequency of this class:

i	x_i	Class	Lower Boundary	Upper Boundary	Observed class Frequency
1 to 10	18.1588 to 19.9170	1	18.000	20.0000	10
11 to 20	20.0147 to 21.9187	2	20.0000	22.0000	10
21 to 30	22.1891 to 23.8770	3	22.0000	24.0000	10
31 to 40	24.0179 to 25.9695	4	24.0000	26.0000	10
41 to 50	26.0840 to 27.9914	5	26.0000	28.0000	10
51 to 60	28.0811 to 29.8964	6	28.0000	30.0000	10
61 to 70	30.1193 to 31.8848	7	30.0000	32.0000	10
71 to 80	32.0863 to 33.9513	8	32.0000	34.0000	10
81 to 90	34.1517 to 35.8416	9	34.0000	36.0000	10
91 to 100	36.0070 to 37.8795	10	36.0000	38.0000	10

Table 6.1 100 simulated values from U[18,38): run 1 of 25

The observed frequencies in all classes are 10, so that the test statistic, $\Sigma(O-E)^2/E$, has the value 0, showing that the shape is acceptable at any level of significance.

Hence for this run the generated output from U[18,38) has been demonstrated to be perfect in respect of the sample mean and standard deviation, and also perfect in terms of the goodness of fit test used here. This verifies the accuracy of the programming of this distribution, and the precision of the generated output.

6.2.3. Modelling the general Triangular distribution T(a,m,b)

In this section the Triangular distribution used is T(10,30,40), so that the least possible, modal and greatest possible values are 10, 30 and 40 respectively.

The expected value is $\mu = (a + m + b) \div 3, = 80 \div 3, = 26.6667$ to 4 d.p.s., and the s.d. is $\sigma = [(a^2 + b^2 + c^2 - b*c - c*a - a*b) \div 18]^{1/2}, = (700 \div 18)^{1/2}$, $= 6.2361$.

The 100 values per run are displayed in the first 100 cells in the respective column of the 25 columns (i.e. runs) in Table 2 in Appendix III. Thus the 100 simulated values in the first run are 12.1841, 12.7143, . . . , 39.8854, and form the contents of cells A1 to A100 in this table. At the base of this column cell A104 contains the sample mean which is calculated via the Excel function AVERAGE(A1..A100), and is 26.6667. Similarly the s.d. of the 100 values is calculated via STDEVPA(A1..A100), and is 6.2361. Hence the sample mean and standard deviation are perfect.

The c.d.f. for this distribution for values of x from 10 to 40 will be in two parts:

$F(x) = (x\text{-}10)^2 \div 600$ if $10 \leq x < 30$, or $F(x) = 1 - (40\text{-}x)^2 \div 300$ if $30 \leq x \leq 40$.

Hence the upper class boundaries of the 10 equi-probable sub-intervals of [10,40) can be found by solving $F(x) = d \div 10$, for $d = 1, 2, \ldots, 10$. For example, when $d = 1$, this upper class boundary is $u_1 = 10 + (60 * 1)^{1\,2}$, $= 17.7460$, and this will then be the lower class boundary of the next class, l_2, say.

Table 6.2 below shows for each class its lower and upper class boundaries, the first and last members observed from the table to lie in this class, and therefore the observed frequency of this class:

i	x_i	Class	Lower Boundary	Upper Boundary	Observed Class Frequency
1 to 10	12.1841 to 17.4353	1	10.000	17.7460	10
11 to 20	17.8760 20.7041	2	17.7460	20.9545	10
21 to 30	21.1913 23.2330	3	20.9545	23.4164	10
31 to 40	23.5668 25.4525	4	23.4164	25.4919	10
41 to 50	25.5825 27.1982	5	25.4919	27.3205	10
51 to 60	27.4559 28.9012	6	27.3205	28.9737	10
61 to 70	29.0111 30.4598	7	28.9737	30.5132	10
71 to 80	30.6367 32.0703	8	30.5132	32.2540	10
81 to 90	32.3491 34.2674	9	32.2540	34.5228	10
91 to 100	34.5712 39.8854	10	34.5228	40.0000	10

Table 6.2 100 simulated values from T(10,30,40): run 1 of 25

The observed frequencies in all classes are 10, so that the test statistic, $\Sigma(O-E)^2/E$, has the value 0, showing that the shape is acceptable at any level of significance.

In 3 runs (number 3, 6 and 16) the observed frequencies are not perfect, but the calculated χ^2 value is still highly non-significant.

Hence for this run the generated output from T(10,30,40) has been demonstrated to be perfect in respect of the sample mean and standard deviation, and also perfect in terms of the goodness of fit test used here. This verifies the accuracy of the programming of this distribution, and the precision of the generated output.

6.2.4. Modelling the general Normal distribution $N(\mu,\sigma^2)$

In this section the Normal distribution used is N(100,36). The expected value is μ = 100 exactly, and the s.d. is $\sigma = 36^{1/2}$, = 6 exactly. The least possible and greatest possible values have been defined within the code to be $\mu \pm 4\sigma$, and so they are 76 and 124 respectively.

The 100 values per run are displayed in the first 100 cells in the respective column of the 25 columns (i.e. runs) in Table 3 in Appendix III. Thus the 100 simulated values in the first run are 85.8737, 86.8429, . . . , 122.4235, and form the contents of cells A1 to A100 in this table. At the base of this column cell A104 contains the sample mean which is calculated via the Excel function AVERAGE(A1..A100), and is 100.0000. Similarly the s.d. of the 100 values is calculated via STDEVPA(A1..A100), and is 6.0000. Hence the sample mean and standard deviation are perfect.

The c.d.f. is not an analytic function and so simulated values have to be generated in conjunction with tabled (or calculated) values of the standard Normal distribution, as defined earlier in Chapter 5.2.

Hence from any table of the percentage points of the standard Normal distribution the class boundaries of the ten equi-probable classes can be calculated. For example, in the first class the lower class boundary will be 100 – 4*6, = 76.000, and the upper class boundary will be 100 – 1.2816*6, = 92.3104.

Table 6.3 below has the same column headers as Table 6.2.

i	x_i	Class	Lower Boundary	Upper Boundary	Observed Class Frequency
1 to 10	85.8737 to 92.1960	1	76.000	92.3104	10
11 to 20	92.5762 to 94.8388	2	92.3104	94.9504	10
21 to 30	95.2093 to 96.7414	3	94.9504	96.8536	10
31 to 40	96.9917 to 98.4284	4	96.8536	98.4802	10
41 to 50	98.5299 to 99.8338	5	98.4802	100.0000	10
51 to 61	100.0511 to 101.4504	6	100.0000	101.5198	11
62 to 71	101.6556 to 103.1271	7	101.5198	103.1464	10
72 to 81	103.1685 to 104.9481	8	103.1464	105.0496	10
82 to 91	105.1257 to 107.4719	9	105.0496	107.6896	10
92 to 100	108.0038 to 122.4335	10	107.6896	124.0000	9

Table 6.3 100 simulated values from $N(100,6^2)$: run 1 of 25

The calculated value of χ^2 is 8 * 0 + 1/11 + 1/9, = 0.2020, and the number of degrees of freedom = 7. On the usual assumption in the null hypothesis that the errors are normally distributed, this result is not even significant at the 99.9 percent level, when the critical value of χ^2 is 0.5549.

Hence for this run the generated output from $N(100,6^2)$ has been demonstrated to be perfect in respect of the sample mean and standard deviation, and also extremely good indeed in terms of the goodness of fit test used here. This verifies the accuracy of the programming of this distribution, and the precision of the generated output.

As before a total of 25 runs were carried out. Runs 3 and 11 are the worst runs. For run 3, the observed frequencies are 10, 11, 9, 11, 9, 10, 10, 9, 10, 11 with χ^2 = 0.6061. Run 11 is very similar, again with χ^2 = 0.6061. Even so this calculated value of X^2 is significantly small, so that even in these worst two runs, the degree of fit is significantly good.

6.2.5. Modelling the general Beta distribution B(a,m,b)

In this section the Beta distribution used is B(12, 15, 30), so that the least possible, modal and greatest possible values are 12, 15 and 30 respectively.

In practice the most useful forms of the Beta distribution occur when the gradients of the two finite tails on both zero, so that the two shape parameters, v and w, are both greater than 2. For convenience, it has been assumed that if the skewness is positive (as here) then v = 3, and otherwise w = 3. Because the mode is a + (b – a)(v – 1)/ (v+w+2), in this example the value of w is 11.

The expected value is $\mu = a + (b - a) * v / (v + w)$, = 15.8571 to 4 d.p.s. with $v = 3$ and $w = 11$, and the s.d. is $\sigma = \left[(\frac{b-a}{v+w}) * \sqrt{(\frac{v*w}{v+w+1})} \right]^{\frac{1}{2}}$ = 1.9070.

The 100 values per run are displayed in the first 100 cells in the respective column of the 25 columns (i.e. runs) in Table 4 in Appendix V. Thus the 100 simulated values in the first run are 12.6418, 12.7258, . . . , 23.0731, and form the contents of cells A1 to A100 in this table. At the base of this column cell A104 contains the sample mean which is calculated via the Excel function AVERAGE(A1..A100), and is 15.8571. Similarly the s.d. of the 100 values is calculated via STDEVPA(A1..A100), and is 1.9070.

Hence the sample mean and standard deviation are perfect.

The c.d.f. is not an analytic function and so simulated values have to be generated in conjunction with calculated values of the standard Beta and Gamma distributions, as defined earlier in Chapter 5.2.

In general, if the variable in question, x, lies in the range $a \leqslant x \leqslant b$ and the Beta distribution is defined in terms of the shape parameters v and w, then the p.d.f. of x is given by $f(x) = \{ (x-a)^{v-1}(b-x)^{w-1} \} / \{ (b-a)^{v+w-1}\beta(v,w) \}$,
where $\beta(v,w) = \Gamma(v)\,\Gamma(w) / \Gamma(v+w)$.

The general Gamma function, $\Gamma(\theta)$, $\equiv \int_{x=0}^{\infty} e^{-x} x^{\theta-1}\,dx$.

If θ is an integer then $\Gamma(\theta) = (\theta-1)!$

The standardised value of x is achieved by the mapping $z = (x-a)/(b-a)$, so that $0 \leqslant z \leqslant 1$.

In this example, $a = 12$, $m = 15$, $b = 30$, $v = 3$, and $w = 11$.
$\therefore \beta(v,w) = \Gamma(3)\,\Gamma(11) / \Gamma(14)$, so that $\beta(3,11) = 2!*10! / 13! = 1/858$,
and $z = (x - 12) / 18$.

The probability that z is at most k is

$$F(z) = \int_{z=0}^{k} f(z)\,dz, = \int_{z=0}^{k} 858\{ z^{-2} (1-z)^{-10}\}dz.$$

Using integration by parts, the upper class boundaries of the 10 equally-probable sub-intervals of B(12,15,30) can be found by de-standardising the solution, z, of the equation:

$$F(z) = 1 - 66(1-z)^{13} + 143(1-z)^{12} - 78(1-z)^{11} = k, \text{ where } k = 0.1, 0.2, \ldots, 0.9.$$

For example, when $k = 0.1$ then solving by the Newton-Raphson method (Celia, 1969) yields $z = 0.0880$ to 4 d.p.s, so that the first upper class boundary is "$x = a + (b-a) * z$", $= 13.5839$ to 4 d.p.s. Consequently in a run of size 100, we'd expect roughly 10 values to lie in the first decile, so that their values would lie between 12 and 13.5839.

Table 6.4 below shows for each class its lower and upper class boundaries, the first and last members observed from the table to lie in this class, and therefore the observed frequency of this class:

i	x_i	Class	Lower Boundary	Upper Boundary	Observed Class Frequency
1 9	12.6418 to 13.5004	1	12.0	13.5839	9
10 20	13.5897 to 14.1560	2	13.5839	14.1679	11
21 30	14.1959 to 14.6434	3	14.1679	14.6619	10
31 40	14.6934 to 15.0991	4	14.6619	15.1314	10
41 50	15.1521 to 15.6004	5	15.1314	15.6081	10
51 60	15.6337 to 16.0817	6	15.6081	16.1190	10
61 70	16.1448 to 16.6469	7	16.1190	16.7004	10
71 80	16.7202 to 17.3738	8	16.7004	17.4205	10
81 91	17.4635 to 18.4619	9	17.4205	18.4760	11
92 100	18.6662 to 23.0731	10	18.4760	30.0000	9

Table 6.4 100 simulated values from B(12, 15, 30): run 1 of 25

The calculated value of χ^2 is 8 * 0 + 1/11 + 1/9, = 0.2020, and the number of degrees of freedom = 7. On the usual assumption in the null hypothesis that the errors are normally distributed, this result is not even significant at the 99.9 percent level (i.e. < 0.5549).

Hence for this run the generated output from B(12, 15, 30) has been demonstrated to be perfect in respect of the sample mean and standard deviation, and also extremely good indeed in terms of the goodness of fit test used here. This verifies the accuracy of the programming of this distribution, and the precision of the generated output.

As before a total of 25 runs were carried out. Seven of these runs were the worst, with the calculated value of χ^2 equal to 0.40404. Again, these worst values were significantly small, showing that the degree of fit is significantly good.

6.3 Assessing the precision of the sampled correlation coefficients

6.3.1. Introduction

This section demonstrates the results and analysis of these results arising from two separate uses of the RCM. In the first example, the two variables both have U[0,1) distributions. In the second example the distributions of the two variables are Triangular and Beta distributions, with respective parameters (10, 30, 40) and (12, 15, 30). From now on these are referred to as T(10,30,40) and B(12,15,30) respectively.

The first example as demonstrated in section 6.3.2. could equally well have used two general Uniform distributions, but the linear transformations make this added generality pointless at this testing stage. In the second example as demonstrated in section 6.3.3., the skewnesses of the two variables are negative and positive respectively.

Numerous combinations of distributions (and skewnesses), sample sizes and target correlations have been generated and examined during this research, and these are listed in Appendix V. A selection of key or representative combinations are reported and analysed in greater detail, either in this chapter or in Appendix VIII. If anything the results reported from this second example, in section 6.3.3., are the least satisfactory of all the combinations tested, and an objective here is to show that, even in this relatively poor set of results, the final results are very good. That is, during the testing stage of the RCM, very many tests were carried out on samples of simulated paired variables, including every combination of variables with large or small positive or negative skewness (where appropriate) and having Normal, Beta, Uniform or Triangular distributions.

As mentioned previously, an objective of this research was not to provide this product moment correlation tool for any continuous distribution, so that the research was ultimately limited to variables with the four above distributions.

For each of the two examples six cases were investigated, in which the target product moment correlation coefficients were assigned values –0.6, -0.4, -0.2, 0.3, 0.5 and 0.7. These six values are meant to be indicative of a broad range of correlation values. It should be noted that, in the cases of perfect negative and positive correlations, -1.0 and 1.0, the RCM easily generates samples whose correlations are indeed –1.0 and 1.0 respectively.

One case is reported for the combination of the two U[0,1) distributions in section 6.3.2, with a further 5 cases in Appendix VIII, and similarly two cases of the combination of T(10,30,40) and B(12,15,30) are examined in section 6.3.3, and the remaining 4 cases are also in Appendix VIII. In each of these 12 cases, the RCM was run with, in turn, sample sizes of 10, 100 and 500 values, thus creating 36 separate scenarios within sections 6.3.2 and 6.3.3 (together with Appendix VIII). In each of these scenarios 25 separate runs of the model were carried out.

For example, in the scenario 'case (1) of T(10,30,40) combined with B(12,15,3), where ρ = -0.6, when sample sizes are 10 pairs)' the "mean = -0.608"' result reports that the mean value of the 25 sample correlation coefficients in these 25 runs is –0.6084 before the swapping procedure described in section 5.6 is applied, and this is improved to –0.5987 as a result of this swapping.

In section 6.3.2, case (1) (with ρ = -0.6) is tabulated, with the other 5 cases in Appendix VIII, and here the swapping process described in section 5.6 is not applied because the two distributions are, essentially, the "building block" U[0,1) distributions. Hence these tables simply record the mean and median values, etc., of the correlation coefficients in the 25 separate runs for each of the three sample size scenarios.

In section 6.3.3 cases (1) and (6) are tabulated (the other 4 cases being again in Appendix VIII), and in these six cases the tables are split into two parts: the first part shows the mean and median values, etc., of the correlation coefficients in the 25 separate runs for each of the three sample size scenarios before the swapping process described in section 5.6, and the second part shows how the precisions are improved as a result of this swapping. Progress was almost always made, so that swapping improved the accuracy of the generated correlation coefficients, usually very significantly on a relative scale.

6.3.2. Tabulation of the generated sample correlation coefficients from paired U[0,1) distributions.

Combining U[0,1) with U[0,1) via the RCM: Case (1): ρ = -0.6 Distribution of the Generated Correlation Coefficients			
Sample Size:	10	100	500
Mean	-0.5994	-0.6000	-0.6000
Median	-0.5997	-0.6000	-0.6000
Std.Dev	0.0028	0.0000	0.0000
Range	0.0144	0.0000	0.0000
Minimum	-0.6070	-0.6000	-0.6000
Maximum	-0.5927	-0.6000	-0.6000

* This table does not distinguish between the figures "Before" and "After" swapping as no swapping is required at this stage to improve the distributions of the generated correlation coefficients for U[0,1) with U[0,1).

6.3.3. Tabulation of the generated sample correlation coefficients from paired T(10,30,40) and B(12,15,30) distributions.

Combining T(10,30,40) with B(12,15,30) via the RCM: Case (1): $\rho = -0.6$

Distribution of the Generated Correlation Coefficients

	Using the RCM *Before* Swapping			Using the RCM *After* Swapping		
Sample Size:	10	100	500	10	100	500
Mean	-0.6084	-0.5977	-0.5992	-0.5987	-0.6000	-0.6000
Median	-0.6097	-0.5988	-0.5992	-0.5995	-0.6000	-0.6000
Std.Dev.	0.0802	0.0117	0.0034	0.0030	0.0000	0.0000
Range	0.2886	0.0477	0.0151	0.0132	0.0000	0.0000
Minimum	-0.7389	-0.6207	-0.6039	-0.6023	-0.6000	-0.6000
Maximum	-0.4503	-0.5730	-0.5888	-0.5891	-0.6000	-0.6000

Combining T(10,30,40) with B(12,15,30) via the RCM: Case (6): $\rho = -0.7$

Distribution of the Generated Correlation Coefficients

	Using the RCM *Before* Swapping			Using the RCM *After* Swapping		
Sample Size:	10	100	500	10	100	500
Mean	0.6667	0.6821	0.6815	0.6989	0.7000	0.6984
Median	0.6700	0.6793	0.6813	0.6997	0.7000	0.6989
Std.Dev	0.0424	0.0077	0.0029	0.0028	0.0000	0.0017
Range	0.1680	0.0268	0.0118	0.0141	0.0002	0.0061
Minimum	0.5794	0.6694	0.6760	0.6880	0.6999	0.6940
Maximum	0.7474	0.6963	0.6878	0.7021	0.7001	0.7001

6.3.4. Analysing the before and after swapping results when applied to two distributions which are Triangular(10,30,40) and Beta(12,15,30)

Unlike the example shown in section 6.3.2., the swapping procedure described in section 5.6 was required when applied to the example in section 6.3.3. One way to measure its effectiveness is to calculate the percentage of the error in the test statistic before swapping which is corrected as a result of swapping. Call this the correction factor. For example, in Case (1) the absolute error in the mean value before swapping is 0.0084, whereas after swapping it reduces to 0.0013, so that the correction factor is (1 – 0.0013 / 0.0084) * 100, = 84.52%.

The following table calculates these correction factors in each of the six cases for the mean value of the 25 paired-samples' calculated correlation coefficients, their median, standard deviation and range.

In one case, the value of the median after swapping in Case (2) in Appendix VIII, the value is actually correct to 4 d.p.s, so that the maximum possible absolute error is then 0.00005, and this figure has been used in the calculation of the correction factor: (1 – 0.00005 / 0.0154) * 100%, = 99.68%.

Statistic	Correlating T(10,30,40) and B(12,15,30): % Correction Factor in each of the Six Cases when the sample size = 10						Average % Correction Factor
	(1)	(2)	(3)	(4)	(5)	(6)	
Mean Value	84.52	95.74	95.98	99.12	99.08	96.67	95.19
Median	94.85	99.68	95.00	99.27	95.80	99.00	97.27
Std Deviation	96.26	97.39	95.97	96.39	95.73	93.40	95.86
Range	95.43	97.36	93.98	96.62	96.29	91.61	95.22

Table 6.5 Correction factor showing the improvement of results after swapping when sample size is 10

Averaging these four average percentage correction factors yields an average improvement arising from swapping of over 96 percent. Even with samples as small as 10, as they are here, only one of the above 24 correction factors is less than 90 percent, so that the contribution made by swapping here is quite impressive.

When the results corresponding to samples of size 100 are analysed, the results after swapping are excellent. Only in case (6) is there any inaccuracy at all when measured to 4 d.p.s, and even in this case the smallest and largest of the resulting correlation coefficients are 0.6999 and 0.7001 when the target sought was 0.7.

Similarly the analysis of the results when the sample sizes are 500 shows that the results after swapping are again excellent except in one case, number (6) again, where a number of the correlation coefficients arising from the generated samples were a little small. The average sample mean was around 0.2 percent below the target (0.7), although the median was a little closer to the target, suggesting that these data are positively skewed. The maximum correlation coefficient was 0.7001, which is fine, but the smallest was only 0.6940, which is an error of 0.86%. However this 'blip' is easily explained.

At present in any run the programmed swapping procedure stops *either* when the correlation coefficient is within 0.00005 of the target *or* after a specific number of rounds of swaps (5), whichever is reached first. This '5' limit is applied irrespective of the sample size, whereas a probabilistic analysis suggests that the permissible maximum number of rounds should be proportional to the square root of the sample size. In practice the swapping procedure with samples of size 10 virtually always took only 3 rounds or less. Hence the square-root rule would suggest that samples of size 500 might require up to around 20 rounds (possibly more) before the approach ceases to make any further progress.

[The speed with which this procedure is carried out on the current range of p.c.s is very fast indeed, so that increasing the number of possible rounds of swapping from 5 to 20 would add just a second or so at most to the run time.]

However, to put this apparently poor result into a proper context, the table in section 6.4.3. below shows that in the 25 corresponding runs carried out by @RISK when the targetted correlation was 0.7, it generated one sample of 500 paired values in which the (rank) correlation coefficient was 0.6333 only, so that even here the minimum value 0.6940 generated by the RCM is still better by a factor of around 11 to 1. Further analysis of this case is detailed in section 6.4.3.

6.4 Comparing the final results with one commercial risk analysis package

6.4.1. Introduction

In this section the final "after swapping" results of the three cases carried out in the two examples, which were tabulated in sections 6.3.2. and 6.3.3., are compared with the corresponding results of runs of @RISK using the same parameter values and sample sizes. The other nine cases in Appendix VIII are similarly compared with @RISK in Appendix IX.

The analysis with respect to Crystal Ball has not been as rigorous, but generated broadly similar results, and so has not been included in this thesis for reasons of desirable brevity.

For example in case (1) in the second table below in section 6.4.3., the first reported value "mean = -0.4708" shows that, when running @RISK 25 times for the two triangularly distributed and beta distributed variables with samples of size 10 and a target correlation coefficient equal to –0.6, the average of the 25 values of the 25 paired-samples' resulting correlation coefficients was -0.4708. These values are, of course, rank correlation coefficients. This figure, -0.4708, can be compared with the corresponding "after swapping" result generated by the RCM which is copied across from case (1) in section 6.3.2. This is -0.5987.

In this example, the relative accuracy of these two averages can be calculated. Thus, the absolute error using @RISK is 0.1292 (= abs(-0.4708 – (-0.6))), whereas the absolute error using the RCM is only 0.0013. Hence, the ratio of these two absolute errors is 0.1292 to 0.0013, or approximately 99.4 to 1. i.e. The RCM is much more accurate in relation to this specific test statistic in this case.

The reported (average) values of the median can be compared similarly. However, the comparison in relation to the other four reported statistics is a little different. For example, in section 6.4.3. in case (1), the two average values of the ranges are 0.7953 for the 25 samples of size 10 generated by @RISK and 0.0132 for the corresponding output from the RCM.

Ideally, of course, the ranges should be zero if perfect precision is attained. This is appropriate in all six cases in section 6.4.2. when the sample size is either 100 or 500, with just one exception. (The exception is case (6) which, as previously mentioned, would almost certainly have been improved if the swapping algorithm was applied for a few more iterations.) So, in this case the ratio of the two absolute errors will be 0.7953 to 0.0132, or 60.25 to 1.

The performance of @RISK and the RCM after swapping, for the six cases of each of these two examples, are tabulated in sections 6.4.2. and 6.4.3. respectively.

Having shown the 12 tables tabulating "@RISK" against "after swapping", a more comprehensive analysis of the results tabulated in sections 6.4.2. and 6.4.3. is assigned to section 6.4.4. In each of the six cases of the two examples we look for the critical values of the absolute error ratio.

6.4.2. Comparing the distribution of generated correlations generated by the RCM and @RISK when pairing U[0,1) and U[0,1).

Combining U[0,1) with U[0,1): Case (1): ρ = -0.6

Distribution of the Generated Correlation Coefficients RCM versus @RISK

	Using @RISK			Using the RCM		
Sample Size:	10	100	500	10	100	500
Mean	-0.8367	-0.8384	-0.8360	-0.5994	-0.6000	-0.6000
Median	-0.8379	-0.8333	-0.8335	-0.5997	-0.6000	-0.6000
Std.Dev	0.0535	0.0150	0.0083	0.0028	0.0000	0.0000
Range	0.1941	0.0491	0.0283	0.0144	0.0000	0.0000
Minimum	-0.9398	-0.8669	-0.8519	-0.6070	-0.6000	-0.6000
Maximum	-0.7457	-0.8178	-0.8236	-0.5926	-0.6000	-0.6000

6.4.3. Comparing the distribution of correlations generated by the RCM and @RISK when pairing T(10,30,40) and B(12,15,30)

Combining T(10,30,40) with B(12,15,30): Case (1): ρ = -0.6

Distribution of the Generated Correlation Coefficients: RCM versus @RISK

Sample Size:	Using @RISK 10	Using @RISK 100	Using @RISK 500	Using the RCM 10	Using the RCM 100	Using the RCM 500
Mean	-0.4708	-0.5953	-0.5818	-0.5987	-0.6000	-0.6000
Median	-0.5328	-0.5991	-0.5827	-0.5995	-0.6000	-0.6000
Std.Dev.	0.2319	0.0591	0.0234	0.0030	0.0000	0.0000
Range	0.7953	0.2161	0.0993	0.0132	0.0000	0.0000
Minimum	-0.8690	-0.7016	-0.6311	-0.6023	-0.6000	-0.6000
Maximum	-0.0737	-0.4855	-0.5318	-0.5891	-0.6000	-0.6000

Combining T(10,30,40) with B(12,15,30): Case (6): ρ = -0.7

Distribution of the Generated Correlation Coefficients: RCM versus @RISK

Sample Size:	Using @RISK 10	Using @RISK 100	Using @RISK 500	Using the RCM 10	Using the RCM 100	Using the RCM 500
Mean	0.6392	0.6710	0.6770	0.6989	0.7000	0.6984
Median	0.6215	0.6713	0.6770	0.6997	0.7000	0.6989
Std.Dev	0.1332	0.0361	0.0180	0.0028	0.0000	0.0017
Range	0.4254	0.1468	0.0815	0.0141	0.0002	0.0061
Minimum	0.4489	0.5824	0.6333	0.6880	0.6999	0.6940
Maximum	0.8743	0.7292	0.7147	0.7021	0.7001	0.7001

6.4.4. Analysing the @RISK and RCM (after swapping) results when applied to two Uniform[0,1) distributions

For each of the two examples in section 6.4.2. and 6.4.3. above, the absolute error ratio (as defined in section 6.4.1.) is calculated across the six cases, for the sample mean, median, standard deviation and range values.

These values are shown in columns (1) to (6) in the two following tables, the first for the results of section 6.4.2., and the second for section 6.4.3. One final preliminary is that, when any of the statistics is correct to the shown accuracy level of 4 d.p.s, the error is taken to be the maximum possible error in such a case, 0.00005.

In the instances where the RCM's results were very accurate the error ratio would clearly be very high indeed. Consequently, the mean ratio might be unduly affected by one or more very high values, so that the median absolute error ratios are reported for the six cases, together with the minimum value.

First the results of correlating samples from the two U[0,1) variables will be investigated. Each value in the columns headed (1) to (6) is the ratio of the absolute error from the @RISK runs to the absolute error using the RCM with swapping.

Sample Size = 10	Correlating U[0,1) and U[0,1): Absolute Error Ratio in each of the Six Cases						Median Abs. Error Ratio	Least Abs. Error Ratio
Statistic	(1)	(2)	(3)	(4)	(5)	(6)		
Mean	394.5	222.8	261.8	29.9	1743.0	458.6	328.2	29.9
Median	793.0	345.0	125.0	175.4	415.8	753.3	380.4	125
Std.Dev.	19.1	52.8	105.1	78.0	30.1	7.9	41.5	7.9
Range	13.5	49.6	77.6	67.6	24.7	5.9	37.2	5.9

Table 6.6 Comparing of @RISK and "after swapping" results when Correlating U[0,1) and U[0,1) with the sample size is 10

6.4.5. Analysing the @RISK and RCM (after swapping) results when applied to two distributions which are Triangular(10,30,40) and Beta(12,15,30)

The table of the ratios of absolute errors when the samples contained 10 pairs of values is as follows:

Sample Size = 10	Correlating T(10,30,40) and B(12,15,30): Absolute Error Ratio in each of the Six Cases						Median Abs. Error Ratio	Least Abs. Error Ratio
Statistic	(1)	(2)	(3)	(4)	(5)	(6)		
Mean	99.4	154.0	131.4	1,225	246.5	55.3	142.7	55.3
Median	134.4	136.0	5.6	1,100	2.3	261.7	135.2	2.3
Std.Dev.	77.3	124.0	119.6	156.0	98.7	47.6	109.1	47.6
Range	60.2	120.6	73.5	196.0	121.7	30.2	97.0	30.2

Table 6.7 Comparing of @RISK and "after swapping" results when Correlating T(10,30,40) and B(12,15,30) the sample size is 10

In the individual cases reported when the sample size is 10, the worst result when benchmarking the performance of the RCM against that of @RISK occurs in case (5), when the target correlation was 0.5. The reported median value using @RISK is 0.4977, so that the error is 0.0023. The corresponding median value of the 25 runs of the RCM is better, and is 0.5010, but the ratio of absolute errors is 0.0023 to 0.0010, or 2.3 to 1. This is the smallest recorded of all the observed improvements of the RCM in comparison with the use of @RISK. The next worst comparison is in case (3), where the target correlation was –0.2, the ratio of absolute errors for the median being 0.0028 to 0.0005, or 5.6 to 1.

In the first of these two cases, however, the sample mean generated by @RISK is 0.4507, compared with 0.5002 for the RCM, and so that their error ratio is 246.5 to 1. The error ratio for the range is 121.7 to 1, and the error ratio for the standard deviation is 98.7 to 1, so that the wide diversity of the values generated by @RISK made it quite lucky that the median value was so close to the target. In the second of these two cases, where the error ratio of the median was only 5.6 to 1, the error ratios of the mean, the range and the standard deviation are 131.4 to 1, 73.5 to 1, and 119.6 to 1 respectively, so that the "5.6" figure is again fortunate, perhaps.

Overall the analysis of the six cases in this section show that in this case the RCM consistently generated samples whose correlation was so much more precise than that generated by @RISK.

Summary

This chapter has demonstrated and tested the accuracy of the results generated from using the RCM in different senarios. Samples of continuous variables on interval scales whose p.d.f.s are any of four different general types have been shown to have perfect means and standard deviations, and their frequency distributions have been shown by chi-square testing to be excellent. This concluded that these samples have passed the first test of being truly representative of the required probability distribution.

Two major test areas in this chapter were, firstly, between the RCM 'before swapping' and 'after swapping'; and, secondly, between @RISK and the RCM after swapping.

When these generated values were used as input to the correlation routines, the closeness of the generated sample product-moment correlation coefficient values to the targetted values was extremely good, and in every instance examined generated correlations which were markedly more accurate than those produced using the possibly inappropriate rank correlation approaches incorporated in popular commercial risk analysis and simulation software such as @RISK.

Numerous other combinations of distributions and skewnesses have been examined during the course of this research, but, in order to avoid a massive dissertation report, only key or representative tests are reported.

Thus it can confidently concluded that the objective of the RCM defined in chapter 1 has been successfully achieved.

The next chapter, Chapter 7, presents a summary of the conclusions drawn in the preceding six chapters; in addition some new conclusions about the contribution of the work are described. A critique of the research is presented, and a number of detailed suggestions are made for further work leading from the research described in this thesis.

Chapter 7: Conclusion, Review, and Recommended Future Work

7.1 Introduction

The objectives of this chapter are five-fold:

- To restate the aim and objectives defined in Chapter 1.
- To examine how closely the research output meets the initial aim and objectives.
- To identify the contribution from this research.
- To review the output from the research.
- To suggest possible extensions of this research.

7.2 Restating the research aim and objectives

The literature review highlighted that a potentially crucial error when using simulation in risk analysis models is to ignore dependency between variables, since this could lead to infeasible or highly unlikely outcomes which would not be identified as such by the simulation.

It thus could waste resources, e.g. money, time and effort, or even lead to inappropriate decisions being made by planners. Any simulation model in a quantitative risk analysis (QRA), therefore, should be monitored closely to prevent it from producing, in any iteration, a scenario that could not sensibly occur (Vose, 2000).

Hence this thesis explores the need to recognise and represent accurately the interdependencies between uncertain components in a model as such.

During the early stages of this research it became very clear that commercial simulation or risk analysis packages, such as @RISK and Crystal Ball, enable some degree of correlation modelling to take place. However, the correlation modelling is based on Spearman's rank correlation technique, which is appropriate for variables represented on an ordinal scale. Where variables are continuous it would be rather more appropriate to measure dependency using Pearson's product moment correlation coefficient. Research has failed to identify any implementations of Pearson's method in commercial simulation or risk analysis packages. There is evidence that bespoke or consultancy packages are increasingly attempting to include product-moment representations, but the evidence seems to be that these are very complex, slow to run, and are all too often not robust.

Filling the gap between acknowledging the importance of modelling correlation and the actual specification and implementation of a procedure for modelling accurate measures of Pearson's correlation has turned into the main aim of this research. To avoid mistakes arising from complexity it was decided to limit the scope of this research to modelling product moment correlations between pairs of variables only.

More specifically:

"To specify, formulate and develop a Pearson product moment correlation model between a pair of continuous variables which can be incorporated into simulation models of complex applications."

Two principal objectives were stated in Chapter 1:

Research objective 1:
The correlation model must generate samples of pairs of values of continuous variables whose Pearson correlation coefficient has acceptable precision

Research objective 2:
The correlation model must include a good representation of the uncertain variables

Comment on these two objectives

Even when commercial packages such as @RISK and Crystal Ball generate rank correlation coefficients their values are often very different from the target values. When the number of variables exceeds 2 this may not be surprising if the user has specified infeasible combinations of partial correlation coefficients (see later).

However, when there are only two variables and the sample size is large, the sample correlation desirably should be close in value to that of the target value. Yet the analysis of the sampled rank correlation coefficients arising from the output of runs of @RISK using samples of size 500 in sections 6.4.2 and 6.4.3 show that some of the results are not at all satisfactory. For example when seeking samples of size 500 with a correlation of 0.3 between two U[0,1) variables, the minimum value achieved in 25 runs was 0.3012 and the maximum was 0.4359, so that every run generated excess correlation. By way of comparison, here, the RCM generated 25 runs, in every one of which the correlation attained was equal to the target value 0.3 to 4 decimal places, and, of course, these measures were Pearson correlation coefficients.

Using Latin hypercubes as a powerful means of variance reduction should enable the sample means and variances to be very close to their expected values. This is true of output from @RISK and Crystal Ball. However, the sample means were not exactly equal to expected values here, and similarly the sample variances differed from their expected values.

Yet in the RCM a simple linear transformation has enabled exact parity to be achieved, without compromising the shape of the distribution of the generated values.

7.3 Meeting the research aim and objectives

The results quoted in the relevant appendices and their analysis detailed in Chapter 6 show conclusively that the general aim has been achieved, and that the two principal research objectives restated above have been met.

Other objectives have also been attained. They are:

- Defining the terminology used in QRA, such as uncertainty, variability, risk, etc. (in Chapter 2).
- Presenting, comparing and contrasting different approaches used in quantifying uncertainty. This formed the basis for the appreciation of simulation (in Chapter 2).
- Identifying how simulation works, together with its advantages over other approaches and its limitations (in Chapter 2).

- Explaining the importance of assessing and including the interdependencies between uncertain variables in a simulation model. This will make possible the construction of a model which allows the interdependencies to be considered and incorporated, through product-moment correlations (in Chapter 2).

- Illustrating how modelling dependencies can be achieved. Throughout the process, various statistical concepts were discussed and it was shown how they can be practically applied (in Chapters 3 and 4).

- Indicating how the RCM can be incorporated into a QRA model in practice. It demonstrates the value and effort of the modelling process (in Chapters 5 and 6).

7.4 Identifying the contribution of this research

This thesis has demonstrated, principally in Chapter 6, but also in Appendices V, VI and VII, that the distributions and sampling of individual continuous variables having assumed probability distributions has indeed been modelled very accurately.

Contributions to knowledge are as follows:

Statistical Contributions:

- Reinforcing the importance of modelling the interdependencies between uncertain components when simulation models are used.

- Discovering a distinctive way of formulating Pearson correlated sample values of a pair of continuous variables during the sampling processes.

- Helping to fill the gap between theoretical awareness of the significance of correlation and the actual practice of its use. In particular recognising the inappropriate use of rank correlations in many situations where the variables are cardinal (i.e. continuous) and developing instead models of product-moment correlations.

Operational Research Contributions:

- Consolidating the advantages and limitations of choosing simulation as a means of carrying out quantitative risk analysis.

- Demonstrating how to improve significantly the reliability and precision of simulation output and exemplifying the sensitivity and confidence of using simulation methodologies.

- Showing the relevance and suitability of simulation and encouraging its wider use.

7.5 Review of the output from this research

During the testing stage of the RCM, very many tests were carried out on samples of simulated paired variables, including every combination of variables with a range of large or small positive or negative skewness (where appropriate) and having general Normal, Beta, Uniform or Triangular distributions. These results are illustrated in Appendix V.

Here the means, standard deviations, and variances of the generated samples are so accurate that they exactly matched the corresponding population values and, therefore, no further explanation is required.

It was not an objective of this research to provide this product moment correlation tool for all continuous distributions, so that the research was ultimately limited to variables with the four above distributions. Clearly it would be easy to extend this modelling to variables with other continuous distributions, such as exponential or Lognormal distributions.

7.5.1. Proof of goodness of fit

The χ^2 test was first used to test the accuracy of the generated distributions against the assumed probability distribution. The results are detailed in Appendix VI which was produced using 100 generated values as per 25 runs to a corresponding probability distribution. This output is summarised in the table below.

	$\chi^2 = 0.0000$	$\chi^2 = 0.2020$	$\chi^2 = 0.4040$	$\chi^2 = 0.6061$
U [18,38)	25	0	0	0
T (10,30,40)	22	3	0	0
N (100,36)	9	11	3	2
B (12,15,30)	9	9	7	0

Table 7.1 Number of runs in 25 attaining the shown levels of χ^2

We can observe that:

- The generated output from U[18, 38) has a perfect goodness of fit test.

- In 3 runs the observed frequencies from T(10, 30, 40) are not perfect, but the calculated χ^2 value is still highly non-significant

- The worst-case results when testing the Beta distribution were still better than the two worst-case results from the Normal distribution. In this latter case, the critical value of chi-square at the extreme 99.5 percent level of significance with 7 degrees of freedom is 0.989 so that even here there's a chance of well under 1 in 200 that the generated sample does not come from a Normal population. However, even in these two worst cases each of the 6 critical observations (out of 100) missed the desired class by very small amounts indeed.

Even here in the two worst cases we may conclude that the degree of fit is significantly good, confirming that sampling from these four distributions using Latin hypercubes and then scaling the sample means and s.d.s has generated samples which attain very high levels of precision and are thus acceptably representative on virtually any scale.

7.5.2. Quantifying the improvement in the correlation coefficient

The successful modelling of the RCM has been testified in two separate ways.

1) Comparing the 'before' and 'after' swapping results when applied to two distributions which are (i) both Uniform [0, 1); and (ii) Triangular(10, 30, 40) and Beta(12, 15, 30).

In case (i) the process of swapping was often never needed in the individual runs (because the results before swapping were already within the accepted tolerance limits). Sometimes when swapping was attempted, no further progress was made.

In case (ii) the effectiveness was measured by calculating the percentage of the error in the test statistic 'before' swapping which was corrected as a result of swapping. This was called the correction factor. It was shown that an average improvement arising from swapping was over 96 percent. Even with samples as small as 10 only one of the 24 "worst case" correction factors documented in section 6.3.4 was less than 90 percent, with the average correction factor being over 96 percent overall, so that the contribution made by swapping here is quite impressive. With larger samples the results were even more impressive.

2) Comparing the 'after swapping' results of the RCM with corresponding runs of @Risk when applied to distributions which are (iii) both Uniform[0,1); and (ii) Triangular(10, 30, 40) and Beta(12, 15, 30).

   Overall the analysis in section 6.4.5. shows that in the case when the sample size is 10, the RCM consistently generated samples whose correlation was so much more precise than that generated by @RISK.

   The smallest of all the observed ratios of improvements of the RCM in comparison with the use of @RISK was 2.3:1, and occurred in case (5) in section 6.4.5. In this case the medians were being compared. Most other improvement ratios were very much better than this with, for example, the median ratio of improvements for the four key parameters tabulated in this section ranging from 97.0:1 to 142.7:1.

The conclusion is that this broadly 100 to 1 improvement in the error ratios demonstrates clearly that the RCM has generated measures of correlation which are highly more accurate than that produced by @RISK. Additionally, of course, Pearson's product moment correlations are generated by the RCM, not by Spearman's rank correlation coefficients. Again it should be noted here that, of all the combinations of variables with different distributions whose results are outlined in Appendix V, the combinations of the distributions which formed the basis of the analysis in Chapter 6 were the ones which were most favourable to the output from @RISK.

7.6 Recommended further work

This research has been initially aimed at generating Pearson's correlation measure for a pair of continuous variables. Separate research is currently generating very promising results when Pearson's partial correlation coefficients are modelled in the case of three continuous variables. Eventually this should be extended to deal with a general number of continuous variables. Even with three variables there are problems with feasibility. For example, suppose the three variables are X, Y and Z, and that the user specifies high positive correlations of 0.9 and 0.8 between X and Y and between X and Z respectively.

Then an eigen value analysis, for example, will indicate that the correlation between Y and Z could not be as small as -0.5, say. This is already recognised in Crystal Ball, for example. An initial paper on this is scheduled for early 2006. Other factors relating to this are discussed in section 2.10 earlier.

A second major area of research here will be to seek means of improving the shape of the individual generated frequency distributions while, at the same time, maintaining the perfect sample means and variances.

Initial research here is also progressing very well, with a perturbation of triplets of the sampled values from an individual distribution indicating a way ahead, so that measures of skewness and/or kurtosis can be improved.

An initial paper on this is scheduled for mid-2006.

The only correlation analysis generated by products such as @RISK relates to rank correlation. Current research into developing a heuristic swapping routine to be used to improve sample rank correlations is going very well, and a research paper on this is at a draft state. The objective here is to formalise this into a research paper by the end of 2005. The experimental work has shown that quite substantial improvements can be made when the number of variables is just two, and it looks very promising when this is increased to four or five.

Ultimately, of course this should be extended to a larger dimension so that it could be considered as an alternative to the Cario-Nelson NORTA-based approaches. This would begin to address the concluding concern of Schmeiser (1999) that " ... the state of the art (simulation) is far from allowing novice practitioners to build complex input models in the way that they can build complex logical models in today's commercial software."

The "rank correlation heuristic" which forms the basis of the research described in the previous paragraph could be ported to become a input model post-processor if the input model has been, for example, the one generated by products such as @RISK or if, more generally, the input model is based on any derivative of the Iman and Conover distribution-free approach. This would alleviate the observations of a number of authors (including Iman and Conover themselves) that the accuracy of the generated rank correlation matrix is too often compromised to a lesser or greater extent.

A possibly even more attractive outcome here would be to port the input rank correlation matrix generated by, for example, @RISK into a future generalised form of the product-moment heuristic swapping algorithm which has been developed in this research and detailed in this thesis. This would enable the practitioner to generate samples with accurate marginal distributions and acceptable product-moment representations of the pair-wise relationships between the sampled variables, without the shear complexity, demand on computing resources, or risk of infeasibility currently posed by the ORTA derivatives. In fact a paper on this by Pryor and Sim was presented at OR47 in September 2005 and, following favourable feedback, will shortly be extended to a working paper prior to submission for publication.

One of the limitations of products such as Microsoft Excel that have been identified during this research has been that these spreadsheet programmes do not have built-in routines to generate rank correlations, and yet if @RISK output including rank correlations is imported into Excel it is then not possible to carry out further post-analysis of this part of the output. Similarly, Crystal Ball (which is marketed as "*an Excel add-in*" can generate only rank correlations, which can the neither be verified nor processed within the spreadsheet.

One way around this is to use an alternative package such as SPSS, which would perhaps pose a needless additional expense to users. There are available a number of spreadsheet rank correlation add-ons which have been produced by commercial companies but, again, these are costly, so that the generation of a simple academic routine to provide rank correlation (and possibly other non-parametric) functions would seem desirable.

Price (2002) has highlighted the convenience of being able to use three-point parameterisations of certain distributions. In his case this was in the context of developing a commercial schedule risk analysis product utilising the Lurie-Goldberg algorithm. For example, the beta distribution requires four parameters to specify it, such as the least and greatest value, together with either the two shape parameters v and w or the modal and mean values μ and m. Within this thesis, in chapter 6, has been described a method of using only three parameters to obtain an acceptable representation of any beta distribution of the category that appears so often in practice. I.e. Where the tails have zero gradients. This part of the thesis is currently being extended into a research paper, and the first draft version has been produced in May 2005."

Finally, the 'triplet perturbation' approach which is proving very promising in the second potential area outlined above could also be effective in improving the Pearson correlation modelling algorithm developed in the RCM.

It is interesting to see the current effect of the RCM on skewness and kurtosis. An initial test of comparing the skewness and kurtosis of generated samples from examples of these distributions has been carried out. The detail of these test results is reported in Appendix VII.

The analysis is summarised in two tables below.

	Sample size = 10	Sample size = 100	Sample size = 500
Uniform[18,38)	0.0174	0.0001	0.0000
Triangular(10,15,40)	0.0849	0.0021	0.0002
Triangular(10,30,40)	0.0959	0.0031	0.0002
Normal(100,36)	0.1387	0.0232	0.0055
Beta(12,15,30)	0.2127	0.0310	0.0098
Beta(12,25,30)	0.1149	0.0196	0.0036

Table 7.2 The maximum absolute difference in the measure of skewness

	Sample size = 10	Sample size = 100	Sample size = 500
Uniform[18,38)	0.0204	0.0001	0.0000
Triangular(10,15,40)	0.2319	0.0059	0.0005
Triangular(10,30,40)	0.1839	0.0078	0.0006
Normal(100,36)	0.6248	0.0922	0.0212
Beta(12,15,30)	0.8301	0.1427	0.0502
Beta(12,25,30)	0.4367	0.0682	0.0148

Table 7.3 The maximum absolute difference in the measure of kurtosis

Generally the accuracy of the modelling of both skewness and kurtosis is seen from these two tables to improve markedly with sample size, as would be expected. At first sight the precision levels of the calculated skewness values seem to be much better than those of the kurtosis values. However, each expected kurtosis value is rather larger than the corresponding absolute size of the expected skewness value so that the relative errors in the calculated kurtosis values are on average only a little worse than those of the calculated skewness values.

The largest single difference in the above two tables is 0.8301 which is when the sample size is 10 in one run of the simulation of values of a B(12,15,30) variable. The research currently being carried out elsewhere into improving skewness and/or kurtosis suggests that this maximum difference can be dramatically improved.

Overall, given that there was no objective in this research to generate highly accurate measures of shape, the above two tables demonstrate that the RCM seems to be dealing acceptably with skewness and kurtosis. Additional non-comprehensive experiments suggest that the measures of skewness and kurtosis currently being generated by the RCM are rather better than the corresponding output from @RISK, but it isn't appropriate to attempt to quantify this improvement here.

In conclusion here the heuristic product-moment correlation improvement algorithm developed within this research has been built on the ideal of generating sample values from the individual (i.e. marginal) distributions using Latin hypercubes. It is to be stressed that this heuristic is equally applicable if the sampling from the marginal distributions is carried out using a Monte Carlo approach, with or without any other form of variance reduction, or a more basic deterministic approach, such as sampling values from a lattice of points.

Summary

It is concluded that, based on the analysis of results generated from this research, the aim of specifying, formulating and developing a Pearson product moment RCM between a pair of continuous variables which can be incorporated into simulation models of complex applications has been achieved successfully.

The extensions identified in the previous section make it possible to envisage a much less complex (and more robust) alternative to the various Cario-Nelson NORTA derivatives when generating samples from a multivariate distribution whose marginal distributions and product-moment (or rank) pair-wise correlations are assumed. Similarly one can envisage an add-on to typical Iman-Conover distribution-free procedures, the benefit of which would be to improve the quite inaccurate representations of the generated rank correlations frequently currently generated by these procedures and/or to generate acceptably accurate product-moment correlations.

The contributions from this research can be seen in two areas. Both the statistical contribution which has discovered a distinctive way of formulating Pearson correlated sample values of a pair of continuous variables during sampling processes, and the operational research contribution which is to improve the reliability and precision of simulation output and exemplify the sensitivity and confidence of using simulation methodologies.

Appendix I

100 generated values as per 25 runs of U[18,38)

Appendix I: U[18, 38)

100 generated values as per 25 runs in column A to Y

	A	B	C	D	E	F	G	H	I
1	18.1588	18.0122	18.1753	18.0945	18.0015	18.0422	18.0001	18.0956	18.1655
2	18.2414	18.3886	18.2246	18.3058	18.3989	18.3586	18.3997	18.3046	18.2348
3	18.5716	18.4306	18.4890	18.4304	18.4859	18.5617	18.5582	18.4869	18.4141
4	18.6286	18.7701	18.7108	18.7699	18.7144	18.6390	18.6416	18.7133	18.7862
5	18.8187	18.8635	18.9131	18.8705	18.8497	18.9073	18.9745	18.8207	18.9045
6	19.1814	19.1372	19.0867	19.1298	19.1506	19.0934	19.0254	19.1795	19.0957
7	19.3084	19.2875	19.2323	19.3943	19.3947	19.3451	19.3411	19.2840	19.2120
8	19.4917	19.5132	19.5676	19.4060	19.4056	19.4555	19.4587	19.5162	19.5882
9	19.6831	19.6850	19.6377	19.7284	19.7883	19.6451	19.7326	19.7374	19.6904
10	19.9170	19.9156	19.9621	19.8719	19.8120	19.9555	19.8673	19.8628	19.9098
11	20.0147	20.1716	20.1203	20.0969	20.1088	20.1918	20.0716	20.1345	20.0397
12	20.3854	20.2290	20.2796	20.3034	20.2915	20.2088	20.3283	20.2657	20.3605
13	20.4903	20.4115	20.4136	20.4165	20.4806	20.5488	20.5522	20.4091	20.4754
14	20.7098	20.7890	20.7862	20.7838	20.7197	20.6518	20.6477	20.7911	20.7248
15	20.9398	20.9797	20.9983	20.9701	20.8528	20.8741	20.9052	20.9747	20.8048
16	21.0603	21.0208	21.0016	21.0301	21.1474	21.1265	21.0947	21.0255	21.1954
17	21.2999	21.2482	21.3875	21.2091	21.2107	21.3467	21.3991	21.3470	21.2449
18	21.5002	21.5523	21.4123	21.5912	21.5895	21.4538	21.4008	21.4532	21.5553
19	21.6814	21.6108	21.7778	21.6800	21.7008	21.6706	21.6462	21.7106	21.6278
20	21.9187	21.9897	21.8221	21.9202	21.8994	21.9299	21.9537	21.8896	21.9724
21	22.1891	22.1563	22.1704	22.0412	22.1769	22.0107	22.1561	22.1160	22.0070
22	22.2110	22.2441	22.2295	22.3590	22.2233	22.3898	22.2438	22.2842	22.3931
23	22.5970	22.4076	22.5900	22.4229	22.5970	22.4428	22.5603	22.5985	22.5727
24	22.6031	22.7929	22.6100	22.7773	22.6032	22.7577	22.6396	22.6016	22.6275
25	22.9183	22.8275	22.8767	22.9189	22.9290	22.8268	22.8840	22.8923	22.9029
26	23.0818	23.1729	23.1232	23.0813	23.0711	23.1736	23.1159	23.1078	23.0973
27	23.3064	23.3340	23.3681	23.3952	23.3431	23.3210	23.2914	23.2739	23.2184
28	23.4936	23.4663	23.4318	23.4050	23.4570	23.4793	23.5085	23.5262	23.5817
29	23.7231	23.6527	23.6766	23.6112	23.7072	23.6365	23.7710	23.6776	23.7910
30	23.8770	23.9476	23.9233	23.9889	23.8929	23.9638	23.8289	23.9225	23.8091
31	24.0179	24.1686	24.0904	24.0159	24.0834	24.0828	24.1141	24.0088	24.0029
32	24.3822	24.2317	24.3095	24.3843	24.3167	24.3175	24.2859	24.3913	24.3972
33	24.4369	24.4422	24.5683	24.4811	24.4868	24.4029	24.4766	24.4260	24.5153
34	24.7631	24.7581	24.6316	24.7190	24.7133	24.7974	24.7233	24.7741	24.6848
35	24.8567	24.8454	24.9642	24.9353	24.8681	24.9530	24.8016	24.8294	24.9216
36	25.1433	25.1549	25.0358	25.0648	25.1320	25.0472	25.1983	25.1707	25.0785
37	25.2038	25.3013	25.3255	25.3603	25.3874	25.3092	25.2260	25.3915	25.3100
38	25.5963	25.4989	25.4744	25.4398	25.4127	25.4910	25.5739	25.4086	25.4901
39	25.6305	25.7240	25.6822	25.7098	25.7274	25.6082	25.6425	25.7754	25.6102
40	25.9695	25.8762	25.9177	25.8903	25.8727	25.9920	25.9575	25.8246	25.9898
41	26.0840	26.1214	26.0274	26.0253	26.1510	26.0458	26.0619	26.0506	26.0503
42	26.3160	26.2787	26.3726	26.3748	26.2491	26.3543	26.3381	26.3495	26.3497
43	26.5097	26.4201	26.5339	26.4739	26.5369	26.4164	26.5849	26.4315	26.4951
44	26.6903	26.7800	26.6661	26.7261	26.6631	26.7837	26.6150	26.7685	26.7049
45	26.9117	26.9886	26.8263	26.8803	26.8459	26.9915	26.9094	26.8070	26.9030
46	27.0883	27.0114	27.1737	27.1198	27.1541	27.0086	27.0906	27.1931	27.0971
47	27.3562	27.2766	27.2005	27.3918	27.2059	27.3930	27.3426	27.3132	27.3401
48	27.4438	27.5235	27.5995	27.4083	27.5942	27.4070	27.4573	27.4869	27.4600
49	27.6086	27.6695	27.6611	27.7032	27.6029	27.6454	27.7606	27.7347	27.7842
50	27.9914	27.9305	27.9389	27.8968	27.9971	27.9546	27.8394	27.8653	27.8158
51	28.0811	28.0080	28.1178	28.1184	28.1683	28.0043	28.1410	28.1372	28.1728
52	28.3188	28.3920	28.2822	28.2816	28.2316	28.3957	28.2591	28.2628	28.2272
53	28.4118	28.5780	28.4696	28.5617	28.4396	28.5166	28.5348	28.4657	28.4810
54	28.7882	28.6219	28.7304	28.6383	28.7604	28.6834	28.6652	28.7343	28.7189
55	28.8624	28.8987	28.9358	28.9527	28.8110	28.8000	28.8362	28.8594	28.8216

Appendix I: U[18, 38)

100 generated values as per 25 runs in column A to Y

	A	B	C	D	E	F	G	H	I
56	29.1376	29.1012	29.0642	29.0473	29.1890	29.1999	29.1638	29.1405	29.1784
57	29.3206	29.2814	29.3098	29.2543	29.2788	29.3261	29.2206	29.3782	29.3418
58	29.4794	29.5185	29.4903	29.5456	29.5211	29.4738	29.5794	29.4217	29.4581
59	29.7035	29.7862	29.6576	29.6129	29.7162	29.7061	29.7890	29.6091	29.7427
60	29.8964	29.8137	29.9425	29.9870	29.8838	29.8938	29.8110	29.9909	29.8573
61	30.1193	30.0502	30.0430	30.0843	30.1191	30.0290	30.1567	30.0923	30.0930
62	30.2807	30.3497	30.3571	30.3156	30.2809	30.3709	30.2433	30.3076	30.3069
63	30.4659	30.5107	30.4879	30.4159	30.4531	30.4586	30.5223	30.5028	30.4710
64	30.7341	30.6891	30.7122	30.7840	30.7468	30.7412	30.6777	30.6972	30.7289
65	30.9818	30.8088	30.8874	30.8213	30.9604	30.8123	30.9569	30.9283	30.9873
66	31.0181	31.1910	31.1127	31.1786	31.0395	31.1874	31.0431	31.0717	31.0126
67	31.3540	31.3627	31.2739	31.2918	31.2978	31.3106	31.2463	31.3685	31.2747
68	31.4460	31.4370	31.5262	31.5081	31.5021	31.4892	31.5537	31.4314	31.5252
69	31.7151	31.6939	31.7968	31.7577	31.7448	31.6748	31.6360	31.6351	31.7410
70	31.8848	31.9059	31.8032	31.8422	31.8551	31.9249	31.9640	31.9648	31.8589
71	32.0863	32.1090	32.0680	32.1150	32.1136	32.1927	32.1808	32.0607	32.0444
72	32.3136	32.2907	32.3321	32.2849	32.2862	32.2070	32.2193	32.3391	32.3554
73	32.5465	32.4752	32.4721	32.4158	32.4154	32.4446	32.5427	32.5445	32.5647
74	32.6535	32.7245	32.7280	32.7840	32.7844	32.7550	32.6574	32.6554	32.6352
75	32.8325	32.8292	32.9947	32.8825	32.8723	32.9590	32.8561	32.8988	32.8995
76	33.1674	33.1704	33.0054	33.1173	33.1275	33.0406	33.1440	33.1011	33.1004
77	33.2119	33.3464	33.2598	33.3120	33.2522	33.3425	33.3557	33.3567	33.2460
78	33.5880	33.4532	33.5403	33.4878	33.5477	33.4571	33.4444	33.4431	33.5538
79	33.6486	33.6439	33.6527	33.7220	33.7486	33.7992	33.6251	33.6082	33.6412
80	33.9513	33.9557	33.9474	33.8778	33.8512	33.8003	33.9750	33.9917	33.9586
81	34.1517	34.0418	34.0469	34.0342	34.0885	34.0448	34.0497	34.0151	34.0884
82	34.2482	34.3577	34.3532	34.3656	34.3112	34.3548	34.3504	34.3848	34.3114
83	34.5902	34.4644	34.5333	34.5588	34.5683	34.5484	34.4655	34.5983	34.4359
84	34.6097	34.7351	34.6668	34.6409	34.6315	34.6511	34.7346	34.6016	34.7640
85	34.8690	34.8845	34.9046	34.9386	34.8676	34.8537	34.9865	34.9746	34.9637
86	35.1309	35.1150	35.0955	35.0611	35.1322	35.1457	35.0136	35.0253	35.0361
87	35.3368	35.3480	35.3867	35.3256	35.2415	35.2760	35.2058	35.2553	35.3569
88	35.4631	35.4515	35.4135	35.4741	35.5582	35.5234	35.5943	35.5445	35.4429
89	35.7582	35.6234	35.6905	35.6228	35.6417	35.6317	35.7389	35.6218	35.7842
90	35.8416	35.9760	35.9096	35.9769	35.9580	35.9677	35.8613	35.9780	35.8156
91	36.0070	36.0868	36.1165	36.0438	36.0624	36.1846	36.1542	36.1835	36.0695
92	36.3928	36.3125	36.2836	36.3559	36.3373	36.2148	36.2460	36.2163	36.3303
93	36.5588	36.5646	36.4633	36.5804	36.5738	36.4037	36.5023	36.4516	36.5500
94	36.6411	36.6348	36.7368	36.6193	36.6259	36.7957	36.6979	36.7482	36.6498
95	36.8966	36.9143	36.8358	36.9593	36.8066	36.9001	36.9635	36.9756	36.9867
96	37.1032	37.0850	37.1644	37.0404	37.1931	37.0992	37.0367	37.0242	37.0131
97	37.2412	37.3534	37.2175	37.2940	37.3449	37.3869	37.2011	37.2565	37.3793
98	37.5586	37.4459	37.5826	37.5057	37.4547	37.4123	37.5991	37.5433	37.4204
99	37.7203	37.6584	37.7627	37.7632	37.6072	37.6852	37.6119	37.7423	37.6806
100	37.8795	37.9408	37.8375	37.8364	37.9924	37.9141	37.9883	37.8574	37.9192
101									
102									
103	mean								
104	28.0000	28.0000	28.0000	28.0000	28.0000	28.0000	28.0000	28.0000	28.0000
105	std.dev								
106	5.7735	5.7735	5.7735	5.7735	5.7735	5.7735	5.7735	5.7735	5.7735
107									

Appendix I: U[18, 38)
100 generated values as per 25 runs in column A to Y

	J	K	L	M	N	O	P	Q	R
1	18.0066	18.0922	18.1320	18.7894	18.1833	18.0012	18.0204	18.0290	18.1022
2	18.3936	18.3081	18.2676	23.6488	18.2167	18.3994	18.3790	18.3708	18.2974
3	18.5413	18.5546	18.4398	34.7925	18.4382	18.5281	18.5405	18.5974	18.5610
4	18.6588	18.6456	18.7599	32.8547	18.7618	18.6724	18.6589	18.6025	18.6386
5	18.8450	18.8709	18.8714	21.1238	18.9722	18.9512	18.9698	18.8293	18.8621
6	19.1552	19.1293	19.1283	24.6107	19.0278	19.0493	19.0296	19.1705	19.1376
7	19.3981	19.2987	19.2994	34.3532	19.2613	19.3574	19.3005	19.2100	19.3143
8	19.4020	19.5015	19.5003	18.9187	19.5387	19.4431	19.4990	19.5898	19.4853
9	19.7838	19.6226	19.6842	36.1905	19.7584	19.6498	19.6903	19.7759	19.7285
10	19.8163	19.9775	19.9155	30.8484	19.8417	19.9506	19.9092	19.8240	19.8711
11	20.1786	20.0160	20.0108	31.5727	20.0617	20.0272	20.0631	20.1417	20.0605
12	20.2215	20.3842	20.3889	19.6054	20.3383	20.3733	20.3364	20.2581	20.3392
13	20.4760	20.5776	20.5256	33.1448	20.5058	20.5684	20.4553	20.4912	20.4786
14	20.7242	20.6226	20.6741	25.3755	20.6942	20.6320	20.7442	20.7087	20.7211
15	20.9260	20.9554	20.8851	37.5505	20.8986	20.8122	20.9347	20.9894	20.9702
16	21.0741	21.0448	21.1146	27.8711	21.1014	21.1882	21.0649	21.0105	21.0295
17	21.3873	21.3721	21.3324	31.8235	21.3820	21.2156	21.3664	21.3925	21.2952
18	21.4128	21.4281	21.4673	29.1703	21.4180	21.5848	21.4332	21.4074	21.5045
19	21.6565	21.6590	21.7010	19.2556	21.6340	21.7881	21.7496	21.6779	21.6085
20	21.9436	21.9412	21.8987	28.0652	21.9661	21.8123	21.8500	21.9220	21.9913
21	22.0267	22.0248	22.0012	37.7262	22.1093	22.1601	22.1033	22.1754	22.1410
22	22.3734	22.3753	22.3986	22.9149	22.2907	22.2402	22.2964	22.2245	22.2588
23	22.5704	22.5348	22.4702	32.1609	22.5770	22.4629	22.5198	22.5089	22.5142
24	22.6297	22.6653	22.7296	33.7945	22.6230	22.7374	22.6799	22.6910	22.6856
25	22.8138	22.9197	22.9925	23.0855	22.9301	22.8353	22.8750	22.9676	22.9200
26	23.1862	23.0804	23.0073	30.0839	23.0699	23.1650	23.1247	23.0323	23.0798
27	23.3475	23.2079	23.3251	26.3838	23.2897	23.2630	23.2827	23.2016	23.2143
28	23.4526	23.5922	23.4748	22.2012	23.5103	23.5372	23.5170	23.5984	23.5855
29	23.6920	23.6520	23.7668	31.2270	23.6912	23.7824	23.7774	23.7073	23.7339
30	23.9081	23.9481	23.8330	25.0842	23.9088	23.8178	23.8223	23.8926	23.8659
31	24.0095	24.0817	24.1108	36.4985	24.0335	24.1039	24.1271	24.0550	24.1550
32	24.3905	24.3183	24.2891	29.2032	24.3665	24.2964	24.2727	24.3449	24.2448
33	24.4414	24.4437	24.5088	29.5966	24.4039	24.4381	24.4148	24.5546	24.5895
34	24.7587	24.7564	24.6911	19.5451	24.7961	24.7621	24.7850	24.6454	24.6104
35	24.8799	24.9912	24.9124	18.4114	24.9881	24.9270	24.8294	24.8156	24.8804
36	25.1202	25.0088	25.0874	29.6020	25.0119	25.0732	25.1705	25.1843	25.1195
37	25.2996	25.2858	25.2462	32.2387	25.2867	25.2228	25.2902	25.3905	25.3624
38	25.5004	25.5142	25.5537	21.5401	25.5133	25.5774	25.5097	25.4094	25.4375
39	25.6225	25.6780	25.7983	26.5814	25.6251	25.7352	25.7819	25.6791	25.6397
40	25.9776	25.9221	25.8016	23.9516	25.9749	25.8649	25.8180	25.9209	25.9602
41	26.1452	26.0112	26.1196	37.2486	26.1453	26.1361	26.1221	26.1114	26.0750
42	26.2549	26.3889	26.2804	31.1513	26.2547	26.2640	26.2778	26.2886	26.3249
43	26.5906	26.4698	26.5110	35.2119	26.4658	26.5060	26.5430	26.5085	26.4554
44	26.6094	26.7302	26.6890	25.7678	26.7342	26.6941	26.6569	26.6915	26.7445
45	26.9510	26.9567	26.8086	37.0976	26.9983	26.9157	26.8931	26.8727	26.9929
46	27.0491	27.0434	27.1913	28.3347	27.0017	27.0844	27.1068	27.1273	27.0071
47	27.3674	27.3953	27.2876	28.8296	27.3064	27.2158	27.3919	27.2008	27.3320
48	27.4326	27.4047	27.5124	33.8050	27.4936	27.5842	27.4081	27.5992	27.4680
49	27.6630	27.7193	27.7726	23.2077	27.7027	27.6652	27.6865	27.6916	27.6980
50	27.9370	27.8807	27.8274	22.4661	27.8973	27.9348	27.9135	27.9084	27.9020
51	28.0287	28.0148	28.0397	24.5896	28.0400	28.0551	28.1175	28.1818	28.1377
52	28.3713	28.3852	28.3603	25.4247	28.3600	28.3449	28.2825	28.2182	28.2624
53	28.5117	28.5637	28.4094	24.3462	28.5868	28.5084	28.4177	28.4610	28.5432
54	28.6882	28.6363	28.7906	30.4910	28.6132	28.6915	28.7823	28.7390	28.6568
55	28.9099	28.8987	28.9751	22.7344	28.8580	28.9665	28.9250	28.8611	28.9402

Appendix I: U[18, 38)

100 generated values as per 25 runs in column A to Y

	J	K	L	M	N	O	P	Q	R
56	29.0901	29.1013	29.0249	27.4667	29.1420	29.0334	29.0750	29.1389	29.0598
57	29.2111	29.2857	29.2910	26.9914	29.3466	29.2754	29.2165	29.2849	29.3457
58	29.5889	29.5142	29.5091	24.9161	29.4534	29.5245	29.5835	29.5151	29.4544
59	29.7081	29.7356	29.7921	18.1995	29.6618	29.6867	29.7895	29.7290	29.6277
60	29.8919	29.8644	29.8080	24.0542	29.9382	29.9132	29.8107	29.8710	29.9723
61	30.1525	30.1906	30.0806	22.1993	30.1113	30.0766	30.0661	30.0879	30.0240
62	30.2474	30.2094	30.3195	34.4069	30.2887	30.3233	30.3340	30.3121	30.3761
63	30.4006	30.4994	30.4272	35.1853	30.5649	30.5232	30.4378	30.4254	30.4957
64	30.7994	30.7006	30.7729	34.8141	30.6351	30.6766	30.7623	30.7747	30.7044
65	30.9496	30.8892	30.9884	28.5244	30.8351	30.8717	30.9554	30.9079	30.9368
66	31.0504	31.1107	31.0117	35.6976	31.1649	31.1282	31.0448	31.0922	31.0633
67	31.3109	31.3585	31.3891	21.2604	31.2192	31.3483	31.3921	31.3632	31.3320
68	31.4891	31.4414	31.4110	20.8768	31.5808	31.4515	31.4081	31.4369	31.4681
69	31.7394	31.6714	31.6716	36.9016	31.6203	31.7895	31.6513	31.7265	31.7284
70	31.8605	31.9285	31.9285	36.7007	31.9797	31.8103	31.9489	31.8736	31.8717
71	32.1038	32.0579	32.1579	20.4150	32.1439	32.1398	32.1533	32.1365	32.1807
72	32.2961	32.3420	32.2423	33.2705	32.2561	32.2599	32.2470	32.2636	32.2195
73	32.4981	32.4603	32.5353	21.7896	32.4297	32.4045	32.4463	32.5443	32.5372
74	32.7019	32.7396	32.6649	30.7088	32.7703	32.7953	32.7539	32.6558	32.6630
75	32.9405	32.9545	32.9825	26.6187	32.8903	32.8512	32.8580	32.9953	32.8157
76	33.0594	33.0454	33.0177	35.9017	33.1097	33.1485	33.1423	33.0048	33.1845
77	33.2746	33.2058	33.2401	32.4471	33.2923	33.2264	33.2883	33.2565	33.2738
78	33.5253	33.5941	33.5601	31.7762	33.5077	33.5733	33.5120	33.5436	33.5265
79	33.6754	33.7377	33.6572	23.5927	33.7225	33.6418	33.7892	33.7179	33.6065
80	33.9245	33.8622	33.9431	27.3333	33.8775	33.9579	33.8111	33.8822	33.9938
81	34.1015	34.0219	34.0739	20.3393	34.1608	34.1103	34.1485	34.1360	34.1253
82	34.2984	34.3780	34.3264	27.0087	34.2392	34.2893	34.2518	34.2641	34.2750
83	34.4249	34.5172	34.5077	37.8730	34.5462	34.5481	34.5721	34.4622	34.5058
84	34.7749	34.6827	34.6926	34.0462	34.6538	34.6515	34.6283	34.7379	34.6945
85	34.9870	34.8251	34.8921	25.8324	34.8260	34.9415	34.9344	34.9796	34.8912
86	35.0129	35.1748	35.1082	20.0613	35.1740	35.0581	35.0661	35.0205	35.1091
87	35.2911	35.2074	35.2692	30.3159	35.2664	35.3286	35.3363	35.2270	35.3416
88	35.5088	35.5924	35.5310	28.6755	35.5336	35.4710	35.4642	35.5731	35.4587
89	35.6456	35.6799	35.7843	19.9953	35.7436	35.6027	35.7026	35.6235	35.6078
90	35.9543	35.9200	35.8160	21.8110	35.8564	35.9969	35.8979	35.9766	35.9925
91	36.0694	36.1914	36.1870	26.0164	36.0045	36.0615	36.1753	36.0050	36.0444
92	36.3304	36.2084	36.2133	33.5291	36.3955	36.3381	36.2252	36.3951	36.3560
93	36.4730	36.5074	36.4729	27.7289	36.4816	36.4798	36.5798	36.5852	36.5670
94	36.7268	36.6924	36.7275	20.7856	36.7184	36.7197	36.6207	36.6149	36.6334
95	36.8242	36.8948	36.9205	36.2087	36.8581	36.9278	36.8860	36.9581	36.8901
96	37.1757	37.1050	37.0798	18.2013	37.1419	37.0717	37.1145	37.0420	37.1103
97	37.3225	37.2638	37.2998	19.0821	37.3431	37.2121	37.2461	37.2366	37.2856
98	37.4774	37.5359	37.5006	35.5875	37.4569	37.5873	37.5545	37.5636	37.5148
99	37.6601	37.7510	37.6830	32.7525	37.7321	37.6616	37.6170	37.7970	37.6918
100	37.9398	37.8488	37.9174	29.9978	37.8679	37.9379	37.9836	37.8032	37.9086
101									
102									
103	mean								
104	28.0000	28.0000	28.0000	28.0000	28.0000	28.0000	28.0000	28.0000	28.0000
105	std.dev								
106	5.7735	5.7735	5.7735	5.7735	5.7735	5.7735	5.7735	5.7735	5.7735
107									

Appendix I: U[18, 38)
100 generated values as per 25 runs in column A to Y

	S	T	U	V	W	X	Y
1	18.0044	18.1804	18.0604	18.0474	18.0094	18.1339	18.0744
2	18.3950	18.2196	18.3406	18.3525	18.3899	18.2660	18.3259
3	18.5441	18.4310	18.4601	18.5596	18.5852	18.4429	18.5495
4	18.6554	18.7690	18.7408	18.6404	18.6141	18.7569	18.6507
5	18.9171	18.8911	18.8795	18.8665	18.9251	18.9375	18.9118
6	19.0824	19.1089	19.1214	19.1335	19.0742	19.0624	19.0885
7	19.3098	19.2627	19.2529	19.2033	19.3910	19.3763	19.3492
8	19.4897	19.5373	19.5480	19.5967	19.4083	19.4235	19.4510
9	19.7337	19.7300	19.7754	19.6580	19.6737	19.7298	19.6019
10	19.8658	19.8700	19.8255	19.9420	19.9257	19.8701	19.9983
11	20.0438	20.0693	20.0671	20.1859	20.1719	20.1469	20.1362
12	20.3558	20.3307	20.3337	20.2141	20.2276	20.2529	20.2641
13	20.4061	20.5654	20.4516	20.4634	20.4766	20.4005	20.4937
14	20.7935	20.6346	20.7492	20.7366	20.7229	20.7994	20.7065
15	20.9294	20.9224	20.8061	20.8350	20.8527	20.9120	20.8279
16	21.0702	21.0776	21.1947	21.1650	21.1468	21.0878	21.1723
17	21.2666	21.3528	21.3563	21.3452	21.3915	21.3981	21.3425
18	21.5330	21.4472	21.4444	21.4547	21.4080	21.4018	21.4577
19	21.7209	21.6449	21.6521	21.6268	21.7616	21.7657	21.6346
20	21.8787	21.9551	21.9485	21.9732	21.8379	21.8342	21.9656
21	22.1981	22.1567	22.1624	22.0829	22.1292	22.1233	22.0877
22	22.2016	22.2433	22.2382	22.3171	22.2704	22.2766	22.3125
23	22.5343	22.4446	22.5252	22.5998	22.4806	22.4570	22.4272
24	22.6653	22.7554	22.6754	22.6002	22.7190	22.7429	22.7729
25	22.9599	22.9484	22.8932	22.8311	22.8018	22.8740	22.8497
26	23.0398	23.0516	23.1073	23.1688	23.1978	23.1259	23.1504
27	23.3678	23.2216	23.2073	23.2787	23.2995	23.2563	23.2461
28	23.4320	23.5784	23.5932	23.5213	23.5002	23.5436	23.5540
29	23.7284	23.7860	23.6478	23.6483	23.6599	23.7468	23.7027
30	23.8713	23.8140	23.9527	23.9517	23.9398	23.8531	23.8974
31	24.1812	24.1283	24.0650	24.1901	24.1059	24.1504	24.0830
32	24.2186	24.2717	24.3354	24.2099	24.2938	24.2496	24.3171
33	24.5563	24.4322	24.5342	24.4505	24.5237	24.5521	24.4293
34	24.6435	24.7678	24.6661	24.7495	24.6761	24.6479	24.7708
35	24.8184	24.8157	24.9508	24.8999	24.9910	24.9972	24.9166
36	25.1815	25.1843	25.0495	25.1001	25.0088	25.0027	25.0835
37	25.3077	25.3170	25.2657	25.3972	25.2650	25.2483	25.3136
38	25.4922	25.4830	25.5346	25.4028	25.5349	25.5517	25.4865
39	25.7009	25.6453	25.7142	25.7558	25.7031	25.7231	25.6333
40	25.8990	25.9547	25.8860	25.8442	25.8968	25.8769	25.9667
41	26.0755	26.0011	26.1262	26.1662	26.1466	26.0523	26.0791
42	26.3244	26.3989	26.2740	26.2338	26.2533	26.3476	26.3210
43	26.4765	26.5992	26.4050	26.5638	26.5301	26.4913	26.5147
44	26.7235	26.6008	26.7951	26.6362	26.6698	26.7086	26.6853
45	26.9371	26.9462	26.8619	26.8220	26.8748	26.8318	26.9265
46	27.0628	27.0538	27.1382	27.1780	27.1251	27.1682	27.0735
47	27.2520	27.2686	27.2983	27.2821	27.2426	27.2258	27.2221
48	27.5479	27.5314	27.5018	27.5179	27.5574	27.5742	27.5779
49	27.7544	27.6489	27.7050	27.6354	27.6712	27.6354	27.6799
50	27.8456	27.9511	27.8950	27.9646	27.9287	27.9646	27.9201
51	28.0936	28.1529	28.1243	28.0619	28.1774	28.1190	28.0603
52	28.3064	28.2471	28.2757	28.3381	28.2226	28.2810	28.3397
53	28.5746	28.4437	28.5853	28.4606	28.5204	28.4685	28.4948
54	28.6255	28.7563	28.6147	28.7394	28.6796	28.7315	28.7052
55	28.9237	28.9501	28.9327	28.8176	28.9946	28.8068	28.9519

Appendix I: U[18, 38)

100 generated values as per 25 runs in column A to Y

	S	T	U	V	W	X	Y
56	29.0764	29.0499	29.0672	29.1824	29.0055	29.1932	29.0480
57	29.2381	29.3322	29.2366	29.3247	29.4000	29.2685	29.2365
58	29.5620	29.4678	29.5633	29.4753	29.4001	29.5316	29.5635
59	29.7974	29.7695	29.6736	29.7570	29.6881	29.7412	29.6022
60	29.8027	29.8305	29.9262	29.8430	29.9120	29.8589	29.9977
61	30.0499	30.0440	30.1538	30.1372	30.1484	30.0879	30.0843
62	30.3503	30.3560	30.2460	30.2628	30.2518	30.3121	30.3157
63	30.5889	30.4489	30.4262	30.5526	30.4815	30.5407	30.5874
64	30.6113	30.7511	30.7735	30.6474	30.7187	30.6593	30.6125
65	30.8201	30.9861	30.8073	30.8757	30.8777	30.9407	30.9675
66	31.1801	31.0139	31.1924	31.1243	31.1226	31.0593	31.0325
67	31.3694	31.2067	31.2563	31.2862	31.2904	31.3528	31.3465
68	31.4308	31.5933	31.5433	31.5138	31.5099	31.4472	31.4534
69	31.6305	31.7865	31.7700	31.6526	31.6868	31.6180	31.6894
70	31.9697	31.8135	31.8296	31.9474	31.9135	31.9821	31.9105
71	32.0075	32.1097	32.1338	32.1470	32.0817	32.1060	32.1081
72	32.3928	32.2903	32.2657	32.2530	32.3186	32.2941	32.2917
73	32.5333	32.5893	32.4647	32.5210	32.4820	32.4291	32.4466
74	32.6670	32.6108	32.7348	32.6790	32.7183	32.7710	32.7532
75	32.9735	32.8903	32.9922	32.9145	32.9133	32.9856	32.8345
76	33.0267	33.1097	33.0072	33.0855	33.0870	33.0145	33.1654
77	33.2374	33.3059	33.2865	33.3249	33.3607	33.2759	33.2984
78	33.5629	33.4941	33.5129	33.4751	33.4397	33.5242	33.5015
79	33.7794	33.6905	33.6599	33.7087	33.6822	33.7133	33.7119
80	33.8210	33.9095	33.9395	33.8914	33.9182	33.8868	33.8880
81	34.0905	34.0802	34.1080	34.1498	34.0645	34.1139	34.1932
82	34.3099	34.3198	34.2913	34.2502	34.3360	34.2862	34.2066
83	34.5341	34.5676	34.4461	34.5371	34.4652	34.4614	34.5712
84	34.6663	34.6324	34.7532	34.6629	34.7353	34.7387	34.6287
85	34.9104	34.8732	34.8219	34.9805	34.8581	34.9089	34.8090
86	35.0900	35.1269	35.1773	35.0195	35.1424	35.0913	35.1908
87	35.3477	35.2408	35.2114	35.2214	35.3819	35.2838	35.2677
88	35.4527	35.5592	35.5878	35.5786	35.4187	35.5163	35.5321
89	35.6749	35.7226	35.6290	35.6230	35.6408	35.6441	35.7124
90	35.9256	35.8774	35.9702	35.9770	35.9598	35.9560	35.8873
91	36.1637	36.0102	36.0054	36.1333	36.1614	36.1951	36.1030
92	36.2368	36.3898	36.3937	36.2667	36.2392	36.2051	36.2968
93	36.5237	36.4861	36.4262	36.5943	36.5452	36.5845	36.5943
94	36.6767	36.7140	36.7729	36.6058	36.6555	36.6157	36.6055
95	36.8517	36.8301	36.8010	36.8231	36.8447	36.8772	36.8190
96	37.1488	37.1699	37.1981	37.1770	37.1560	37.1230	37.1807
97	37.3809	37.3209	37.3022	37.3395	37.3314	37.2537	37.3865
98	37.4196	37.4791	37.4968	37.4606	37.4692	37.5464	37.4132
99	37.6417	37.7672	37.7086	37.6693	37.7028	37.6479	37.7462
100	37.9589	37.8328	37.8904	37.9307	37.8979	37.9523	37.8535
101							
102							
103	mean						
104	28.0000	28.0000	28.0000	28.0000	28.0000	28.0000	28.0000
105	std.dev						
106	5.7735	5.7735	5.7735	5.7735	5.7735	5.7735	5.7735
107							

Appendix II

100 generated values as per 25 runs of T(10,30,40)

Appendix II: Triangular (10, 30, 40)
100 generated values as per 25 runs in column A to Y

	A	B	C	D	E	F	G	H	I
1	12.1841	11.7501	11.9611	12.1535	12.1535	11.5717	11.3582	11.0859	11.9161
2	12.7143	12.9821	12.8060	12.6754	12.6754	13.0642	13.1936	13.3112	12.8486
3	13.5197	13.7870	14.2066	13.8322	13.8322	13.8453	13.7587	13.4976	14.2166
4	14.8782	14.6449	14.2291	14.5827	14.5827	14.5841	14.6847	14.9007	14.2293
5	15.2161	15.0505	15.3294	14.9279	14.9279	14.9392	15.0778	15.3076	15.0565
6	15.7438	15.8642	15.5780	15.9456	15.9456	15.9484	15.8571	15.6689	15.8348
7	16.3749	16.0290	16.2572	16.1008	16.1008	16.4099	16.3375	16.1298	16.2973
8	16.5995	16.8946	16.6580	16.8106	16.8106	16.5331	16.6284	16.8386	16.6275
9	17.2728	17.1962	17.0736	16.9264	16.9264	17.3090	17.1570	17.0225	17.1304
10	17.4353	17.4893	17.5781	17.7210	17.7210	17.3719	17.5421	17.6845	17.5314
11	17.8760	17.8819	17.8353	17.8362	17.8362	18.1079	18.0720	18.0680	17.7655
12	18.3738	18.3514	18.3705	18.3770	18.3770	18.1260	18.1825	18.2033	18.4412
13	18.8260	18.6724	18.7358	18.6038	18.6038	18.7724	18.6714	18.5496	18.4784
14	18.8445	18.9808	18.8963	19.0308	19.0308	18.8782	18.9957	19.1272	19.1473
15	19.4273	19.4457	19.2282	19.3270	19.3270	19.1908	19.1778	19.3204	19.3723
16	19.5508	19.5206	19.7112	19.6225	19.6225	19.7631	19.7921	19.6716	19.5772
17	20.0040	19.9232	19.9964	20.0374	20.0374	20.0802	19.9913	20.0707	19.7985
18	20.1969	20.2661	20.1760	20.1413	20.1413	20.1089	20.2123	20.1489	20.3713
19	20.6509	20.5593	20.5548	20.4009	20.4009	20.6695	20.5140	20.4340	20.6330
20	20.7041	20.7872	20.7746	20.9284	20.9284	20.6760	20.8432	20.9341	20.7017
21	21.1913	21.0552	20.9944	21.1356	21.1356	21.1565	21.0915	21.1090	21.1552
22	21.2576	21.3863	21.4299	21.2969	21.2969	21.2856	21.3621	21.3585	21.2767
23	21.5963	21.5011	21.7272	21.6584	21.6584	21.6783	21.5901	21.6946	21.5534
24	21.8938	21.9829	21.7481	21.8204	21.8204	21.8096	21.9074	21.8178	21.9224
25	22.0289	22.0246	22.1936	22.0531	22.0531	22.0200	22.0124	22.2029	22.2097
26	22.4578	22.4611	22.2836	22.4251	22.4251	22.4655	22.4828	22.3091	22.2706
27	22.7141	22.5224	22.6396	22.5953	22.5953	22.6256	22.5032	22.5087	22.6455
28	22.7357	22.9252	22.8000	22.8469	22.8469	22.8250	22.9532	22.9597	22.7969
29	23.1410	22.9926	22.9699	23.1342	23.1342	23.0092	23.0216	23.1881	23.1101
30	23.2330	23.3817	23.3946	23.2363	23.2363	23.3667	23.3623	23.2095	23.2597
31	23.5668	23.5831	23.5137	23.5967	23.5967	23.6272	23.5087	23.4445	23.6163
32	23.7002	23.6887	23.7494	23.6698	23.6698	23.6466	23.7705	23.8441	23.6499
33	23.9380	23.9115	23.8722	24.0442	24.0442	24.0619	23.9270	23.9192	23.8878
34	24.1927	24.2252	24.2571	24.0898	24.0898	24.0789	24.2178	24.2360	24.2436
35	24.2806	24.3100	24.4162	24.3736	24.3736	24.4294	24.3672	24.3859	24.3576
36	24.6873	24.6668	24.5573	24.6010	24.6010	24.5523	24.6181	24.6098	24.6165
37	24.7247	24.7037	24.8512	24.7457	24.7457	24.8777	24.8824	24.7937	24.7189
38	25.0593	25.0896	24.9406	25.0460	25.0460	24.9216	24.9203	25.0180	25.0721
39	25.1252	25.2350	25.1869	25.1651	25.1651	25.1024	25.2172	25.2359	25.1350
40	25.4525	25.3557	25.4006	25.4230	25.4230	25.4901	25.3797	25.3705	25.4525
41	25.5825	25.6531	25.4911	25.6383	25.6383	25.6439	25.6068	25.6342	25.6316
42	25.7697	25.7127	25.8710	25.7266	25.7266	25.7264	25.7646	25.7463	25.7332
43	25.9426	25.9706	25.9729	26.0347	26.0347	25.9694	25.9509	26.0527	26.0252
44	26.1641	26.1510	26.1484	26.0873	26.0873	26.1572	26.1760	26.0834	26.0969
45	26.2587	26.3940	26.3598	26.2790	26.2790	26.4256	26.3802	26.4164	26.3702
46	26.5845	26.4669	26.5017	26.5813	26.5813	26.4408	26.4857	26.4577	26.4915
47	26.6534	26.7889	26.7683	26.6881	26.6881	26.7433	26.7899	26.6387	26.6641
48	26.9116	26.7947	26.8172	26.8964	26.8964	26.8462	26.7983	26.9557	26.9203
49	27.0735	27.0621	27.0265	27.0262	27.0262	27.0134	27.0155	27.0152	27.0654
50	27.1982	27.2285	27.2667	27.2663	27.2663	27.2830	27.2786	27.2863	27.2279
51	27.4559	27.3314	27.3700	27.3505	27.3505	27.4568	27.3507	27.4905	27.3935
52	27.5076	27.6512	27.6173	27.6355	27.6355	27.5342	27.6360	27.5044	27.5935
53	27.7364	27.7687	27.7691	27.7750	27.7750	27.7569	27.7305	27.7063	27.6722
54	27.9050	27.8949	27.8995	27.8924	27.8924	27.9138	27.9360	27.9664	27.9943
55	28.0193	28.0542	28.0631	28.0018	28.0018	28.0057	28.0378	28.1615	28.1331

Appendix II: Triangular (10, 30, 40)
100 generated values as per 25 runs in column A to Y

	A	B	C	D	E	F	G	H	I
56	28.2871	28.2758	28.2731	28.3319	28.3319	28.3313	28.2949	28.1784	28.2026
57	28.4637	28.3753	28.4223	28.4557	28.4557	28.4123	28.4823	28.3997	28.3925
58	28.4972	28.6095	28.5702	28.5352	28.5352	28.5811	28.5058	28.5938	28.5984
59	28.7015	28.7639	28.7605	28.7403	28.7403	28.7868	28.6776	28.7581	28.6792
60	28.9012	28.8651	28.8766	28.8945	28.8945	28.8508	28.9526	28.8785	28.9556
61	29.0111	29.0369	29.0813	29.1111	29.1111	29.0387	29.0862	29.1204	29.0600
62	29.2229	29.2245	29.1894	29.1574	29.1574	29.2316	29.1772	29.1484	29.2090
63	29.3925	29.3469	29.3156	29.4337	29.4337	29.4364	29.4154	29.3106	29.4002
64	29.4631	29.5369	29.5778	29.4579	29.4579	29.4572	29.4702	29.5790	29.4921
65	29.6146	29.6273	29.7064	29.7034	29.7034	29.6865	29.7007	29.6611	29.7237
66	29.8518	29.8688	29.8013	29.8014	29.8014	29.8199	29.7970	29.8410	29.7819
67	29.9518	30.0316	30.0477	30.0476	30.0476	29.9505	29.9670	29.9334	29.9824
68	30.1183	30.0690	30.0647	30.0617	30.0617	30.1610	30.1347	30.1731	30.1282
69	30.2226	30.2254	30.3275	30.2347	30.2347	30.3435	30.2515	30.2969	30.2835
70	30.4598	30.4896	30.3987	30.4896	30.4896	30.3802	30.4630	30.4209	30.4406
71	30.6367	30.5772	30.6708	30.5616	30.5616	30.6254	30.6173	30.6720	30.6276
72	30.6758	30.7697	30.6890	30.7959	30.7959	30.7318	30.7289	30.6774	30.7297
73	30.9184	30.9355	30.8670	30.9823	30.9823	30.9935	30.9101	30.8896	30.9136
74	31.0481	31.0659	31.1510	31.0295	31.0295	31.0189	31.0912	31.1155	31.1002
75	31.2147	31.2626	31.2278	31.3400	31.3400	31.3028	31.2665	31.2192	31.1999
76	31.4311	31.4188	31.4705	31.3523	31.3523	31.3900	31.4138	31.4652	31.4959
77	31.5659	31.6454	31.6532	31.6537	31.6537	31.5566	31.6610	31.5381	31.5712
78	31.7860	31.7431	31.7523	31.7471	31.7471	31.8464	31.7260	31.8542	31.8326
79	32.0177	31.8995	31.9325	31.8945	31.8945	31.8854	32.0108	31.9406	31.9914
80	32.0703	32.2311	32.2154	32.2499	32.2499	32.2591	32.1149	32.1889	32.1510
81	32.3491	32.3181	32.4513	32.3275	32.3275	32.3137	32.4059	32.3928	32.3183
82	32.5133	32.5865	32.4704	32.5911	32.5911	32.6048	32.4943	32.5094	32.6020
83	32.8241	32.6542	32.8599	32.7088	32.7088	32.8342	32.7039	32.8580	32.7148
84	32.8528	33.0713	32.8802	33.0291	33.0291	32.8993	33.0161	32.8600	33.0243
85	33.1640	33.2448	33.2705	33.0878	33.0878	33.2489	33.0915	33.2196	33.1268
86	33.3796	33.3417	33.3379	33.5212	33.5212	33.3523	33.4975	33.3649	33.4814
87	33.5748	33.7376	33.7450	33.5513	33.5513	33.7575	33.5657	33.7394	33.5725
88	33.8961	33.7753	33.7915	33.9860	33.9860	33.7707	33.9492	33.7700	33.9651
89	34.1988	34.0768	34.0613	34.2018	34.2018	34.1679	34.1805	34.1102	34.1427
90	34.2674	34.4430	34.4858	34.3311	34.3311	34.3639	34.3301	34.4030	34.3938
91	34.5712	34.5867	34.7981	34.7231	34.7231	34.6197	34.6879	34.6621	34.7196
92	34.9986	35.0341	34.8403	34.9091	34.9091	35.0160	34.9208	34.9473	34.9146
93	35.1572	35.4040	35.1519	35.4002	35.4002	35.4248	35.3774	35.1825	35.4219
94	35.6418	35.4347	35.7345	35.4595	35.4595	35.4319	35.4559	35.6616	35.4395
95	35.7960	35.8267	36.0652	35.8165	35.8165	36.0145	35.9994	35.7658	35.8655
96	36.4332	36.4571	36.2265	36.4938	36.4938	36.2673	36.2567	36.5208	36.4384
97	36.6630	36.7237	36.8874	36.9713	36.9713	36.8576	36.5680	36.8929	36.7360
98	37.3190	37.3094	37.1584	37.0585	37.0585	37.1768	37.5026	37.1051	37.3243
99	37.5241	37.7859	38.2343	38.0169	38.0169	38.2053	37.8567	38.1212	38.1418
100	39.8854	38.9583	38.3495	38.6014	38.6014	38.3648	38.8096	38.4196	38.4424
101									
102									
103	mean								
104	26.6667	26.6667	26.6667	26.6667	26.6667	26.6667	26.6667	26.6667	26.6667
105	std.dev								
106	6.2361	6.2361	6.2361	6.2361	6.2361	6.2361	6.2361	6.2361	6.2361
107									

Appendix II: Triangular (10, 30, 40)

100 generated values as per 25 runs in column A to Y

	J	K	L	M	N	O	P	Q	R
1	12.1382	10.7613	12.0200	12.2875	11.9338	11.3891	12.2289	12.2491	12.3019
2	12.7151	13.4387	12.7736	12.5672	12.8383	13.1738	12.6170	12.6071	12.5706
3	13.4747	13.8820	14.0561	13.8817	13.6495	14.1097	13.9458	13.7324	13.5324
4	14.8824	14.6558	14.3810	14.5447	14.7300	14.3725	14.4875	14.6747	14.8360
5	15.0432	15.0412	14.9479	15.0083	15.2112	15.1096	15.0135	14.9835	15.3037
6	15.8698	15.9546	15.9250	15.8809	15.6996	15.8228	15.8750	15.9083	15.6310
7	16.2588	16.5129	16.1439	16.3634	16.0691	16.0689	16.2674	16.1443	16.0807
8	16.6859	16.5213	16.7680	16.5681	16.8397	16.8693	16.6585	16.7798	16.8443
9	17.3096	17.0972	16.9749	17.3226	17.2774	17.0486	17.0040	17.2062	16.9239
10	17.3777	17.6559	17.6736	17.3479	17.3921	17.6381	17.6510	17.4681	17.7373
11	17.9201	18.1150	18.0118	17.9682	17.9720	17.8568	17.9648	17.7478	17.8131
12	18.3140	18.1945	18.2059	18.2525	18.2491	18.3844	18.2548	18.4653	18.4113
13	18.6397	18.7594	18.6498	18.5492	18.6539	18.5411	18.5518	18.7655	18.5873
14	19.0109	18.9598	18.9838	19.0829	18.9841	19.1150	19.0795	18.8797	19.0578
15	19.3287	19.2140	19.3140	19.2303	19.2279	19.2664	19.4357	19.2549	19.4417
16	19.6337	19.8029	19.6323	19.7152	19.7189	19.7042	19.5150	19.6966	19.5201
17	19.9272	19.9057	20.0319	19.9118	19.8332	19.8766	19.9725	20.0772	19.9749
18	20.2604	20.3369	20.1442	20.2635	20.3408	20.3197	20.2039	20.1058	20.2114
19	20.6084	20.6195	20.5169	20.4204	20.5722	20.4347	20.6047	20.6016	20.5503
20	20.7370	20.7783	20.8145	20.9088	20.7642	20.9162	20.7292	20.7371	20.7916
21	20.9978	21.0801	21.0469	21.0571	21.2113	21.0865	21.1709	21.0174	21.1959
22	21.4395	21.4080	21.3818	21.3725	21.2235	21.3640	21.2603	21.4148	21.2435
23	21.6685	21.7175	21.7264	21.6531	21.5858	21.7211	21.7221	21.5972	21.5075
24	21.8176	21.8140	21.7512	21.8239	21.8933	21.7759	21.7552	21.8823	21.9724
25	22.0123	22.2470	22.1288	22.0700	22.0412	22.0190	21.9971	22.1269	22.0304
26	22.4703	22.2815	22.3498	22.4066	22.4385	22.4741	22.4769	22.3543	22.4511
27	22.6175	22.5825	22.5324	22.4988	22.5048	22.5465	22.4871	22.5423	22.6419
28	22.8296	22.9017	22.9069	22.9383	22.9368	22.9093	22.9496	22.8991	22.8039
29	22.9841	22.9966	23.0858	23.1601	23.0214	23.1589	23.1697	23.0344	23.1386
30	23.3870	23.4092	23.2831	23.2077	23.3492	23.2254	23.1982	23.3345	23.2338
31	23.4804	23.5162	23.5064	23.5381	23.6327	23.4443	23.4962	23.5952	23.4766
32	23.7867	23.7830	23.7582	23.7247	23.6359	23.8320	23.7660	23.6705	23.7890
33	23.9126	23.9424	23.8961	24.0094	23.8852	23.9888	23.9841	24.0048	23.9113
34	24.2208	24.2200	24.2352	24.1209	24.2485	24.1560	24.1462	24.1276	24.2208
35	24.3148	24.4670	24.3255	24.3972	24.3297	24.3316	24.4047	24.4780	24.3436
36	24.6587	24.5341	24.6477	24.5739	24.6461	24.6522	24.5668	24.4955	24.6293
37	24.8438	24.8035	24.8585	24.7507	24.8954	24.7602	24.7467	24.8398	24.8803
38	24.9482	25.0108	24.9344	25.0371	24.8999	25.0407	25.0415	24.9507	24.9109
39	25.1400	25.2311	25.1095	25.2763	25.2218	25.2419	25.1643	25.1651	25.2652
40	25.4455	25.3757	25.4773	25.3084	25.3692	25.3549	25.4200	25.4199	25.3210
41	25.5631	25.6177	25.5746	25.5627	25.5357	25.5419	25.5104	25.5610	25.6648
42	25.7979	25.7611	25.7897	25.7967	25.8301	25.8286	25.8486	25.7994	25.6963
43	25.9668	25.9316	25.9406	26.0296	25.9314	25.9290	25.9134	25.9690	25.9825
44	26.1506	26.1999	26.1808	26.0874	26.1919	26.1977	26.2031	26.1484	26.1344
45	26.4246	26.3595	26.2761	26.2552	26.2662	26.3304	26.3093	26.3577	26.3709
46	26.4321	26.5088	26.5846	26.5995	26.5962	26.5354	26.5471	26.4990	26.4851
47	26.7882	26.6631	26.6181	26.6299	26.7259	26.7335	26.7126	26.6474	26.7852
48	26.7910	26.9245	26.9658	26.9482	26.8613	26.8552	26.8676	26.9314	26.7935
49	27.1188	26.9896	27.1115	27.0684	27.0638	27.0778	27.0271	27.0562	27.0697
50	27.1675	27.3023	27.1825	27.2188	27.2315	27.2180	27.2607	27.2310	27.2160
51	27.4849	27.4799	27.4555	27.3659	27.4684	27.4791	27.3564	27.3763	27.4526
52	27.4944	27.5039	27.5325	27.6144	27.5212	27.5099	27.6248	27.6040	27.5265
53	27.6898	27.7668	27.7792	27.6764	27.7483	27.7962	27.7797	27.8058	27.7973
54	27.9680	27.8939	27.8893	27.9836	27.9213	27.8722	27.8826	27.8551	27.8614
55	28.1467	28.0758	28.0720	28.0487	28.1121	28.0397	28.1332	28.1089	28.1253

Appendix II: Triangular (10, 30, 40)
100 generated values as per 25 runs in column A to Y

	J	K	L	M	N	O	P	Q	R
56	28.1789	28.2492	28.2639	28.2793	28.2254	28.2947	28.1967	28.2191	28.2003
57	28.4748	28.4606	28.3413	28.3938	28.3427	28.4002	28.3343	28.4634	28.4826
58	28.5056	28.5172	28.6497	28.5898	28.6494	28.5893	28.6498	28.5199	28.4980
59	28.7741	28.8068	28.8153	28.7886	28.7489	28.6783	28.6741	28.7319	28.7549
60	28.8497	28.8122	28.8214	28.8395	28.8885	28.9540	28.9544	28.8949	28.8689
61	29.0152	28.9767	29.0422	29.0549	29.0760	29.0246	29.0156	29.0137	29.0616
62	29.2405	29.2714	29.2275	29.2061	29.1949	29.2406	29.2464	29.2456	29.1948
63	29.3441	29.3772	29.3543	29.4253	29.3925	29.3245	29.3659	29.3014	29.3941
64	29.5341	29.4920	29.5387	29.4587	29.5016	29.5629	29.5196	29.5803	29.4848
65	29.6368	29.6225	29.7180	29.6324	29.6250	29.7254	29.7462	29.7210	29.7024
66	29.8537	29.8565	29.7888	29.8641	29.8817	29.7751	29.7526	29.7745	29.7891
67	29.9557	29.9559	29.9177	30.0110	29.9516	30.0329	29.9643	30.0312	30.0362
68	30.1396	30.1257	30.1950	30.0904	30.1612	30.0712	30.1394	30.0682	30.0591
69	30.2510	30.3414	30.2999	30.3286	30.2863	30.2598	30.2299	30.2325	30.2422
70	30.4572	30.3500	30.4253	30.3858	30.4396	30.4577	30.4880	30.4814	30.4670
71	30.6146	30.5521	30.6153	30.5599	30.6607	30.5990	30.6154	30.5476	30.5438
72	30.7255	30.7710	30.7436	30.7891	30.6981	30.7508	30.7345	30.7995	30.7986
73	30.9000	30.8392	30.8645	30.9067	30.8423	30.9705	30.8874	30.9830	30.9367
74	31.0957	31.1376	31.1521	31.0973	31.1754	31.0335	31.1191	31.0176	31.0593
75	31.2407	31.1930	31.2299	31.2172	31.2351	31.2682	31.2672	31.2727	31.1914
76	31.4346	31.4606	31.4666	31.4679	31.4614	31.4159	31.4188	31.4084	31.4865
77	31.5378	31.6088	31.5403	31.5358	31.6830	31.6303	31.6296	31.5886	31.6772
78	31.8467	31.7482	31.8662	31.8587	31.7204	31.7610	31.7644	31.8012	31.7056
79	31.9869	31.9744	32.0221	32.0520	31.9650	32.0316	32.0408	31.9924	31.9812
80	32.1343	32.1189	32.1216	32.0788	32.1797	32.0982	32.0924	32.1360	32.1414
81	32.3814	32.3920	32.2756	32.2991	32.3536	32.4162	32.2682	32.4176	32.4242
82	32.5146	32.4731	32.6485	32.6107	32.5672	32.4885	32.6454	32.4857	32.4728
83	32.7703	32.7612	32.8276	32.6609	32.8461	32.6777	32.7033	32.7848	32.7761
84	32.9431	32.9186	32.9103	33.0691	32.8914	33.0485	33.0270	32.9367	32.9388
85	33.1214	33.1010	33.1075	33.2402	33.2428	33.1962	33.1096	33.0930	33.1088
86	33.4619	33.4459	33.5039	33.3519	33.3628	33.3932	33.4899	33.5010	33.4769
87	33.7432	33.6360	33.6524	33.6515	33.7113	33.6739	33.7511	33.6468	33.5440
88	33.7623	33.8312	33.8833	33.8694	33.8222	33.8416	33.7706	33.8702	33.9708
89	34.2449	34.1448	34.1311	34.2523	34.0557	34.2092	34.0094	34.1274	34.1528
90	34.2617	34.3197	34.4083	34.2687	34.4878	34.3065	34.5258	34.3922	34.3577
91	34.7640	34.7430	34.6929	34.5851	34.7726	34.7920	34.7684	34.7108	34.7052
92	34.8399	34.8135	34.9448	35.0439	34.8617	34.8203	34.8537	34.9053	34.9026
93	35.1245	35.1174	35.1571	35.1594	35.3857	35.3714	35.1148	35.1688	35.2825
94	35.7261	35.6792	35.7244	35.7042	35.4777	35.4687	35.7582	35.6884	35.5552
95	35.9830	35.7558	36.0629	36.0132	36.0166	35.9968	35.9083	36.0723	35.9427
96	36.2728	36.4709	36.2242	36.2588	36.2719	36.2668	36.3770	36.1914	36.3200
97	36.8161	36.8567	36.9670	36.8244	36.9503	36.7315	36.9173	36.9081	36.8983
98	37.1924	37.0812	37.0680	37.2033	37.0843	37.3006	37.1046	37.1056	37.1056
99	37.6649	37.5738	38.1170	37.7557	38.1452	37.8655	37.7981	37.7199	37.6680
100	39.2548	39.4255	38.4754	39.0482	38.4406	38.8058	38.9588	39.1284	39.2559
101									
102									
103	mean								
104	26.6667	26.6667	26.6667	26.6667	26.6667	26.6667	26.6667	26.6667	26.6667
105	std.dev								
106	6.2361	6.2361	6.2361	6.2361	6.2361	6.2361	6.2361	6.2361	6.2361
107									

Appendix II: Triangular (10, 30, 40)
100 generated values as per 25 runs in column A to Y

	S	T	U	V	W	X	Y
1	11.6966	11.9866	11.5170	11.9446	12.0267	11.2787	12.0533
2	12.9940	12.7896	13.1158	12.8289	12.7785	13.2307	12.7624
3	13.8792	14.1562	13.7590	14.0016	14.0569	14.2176	13.7428
4	14.5510	14.2809	14.6783	14.4332	14.3896	14.2817	14.6625
5	14.9204	15.1365	15.3731	15.3325	15.0441	15.1206	15.0853
6	15.9595	15.7584	15.5810	15.5833	15.8512	15.8241	15.8169
7	16.4610	16.0440	16.3752	16.3720	16.0280	16.3455	16.0491
8	16.4779	16.8542	16.5860	16.5555	16.8784	16.6248	16.8612
9	17.0684	17.1031	17.1749	16.9659	17.0668	17.0248	17.0205
10	17.5986	17.5524	17.5191	17.6834	17.5955	17.6688	17.6395
11	18.0201	17.9987	17.7545	17.8685	17.8532	17.9477	17.8703
12	18.2085	18.2163	18.4784	18.3449	18.3636	18.3065	18.3486
13	18.5613	18.8033	18.5911	18.7352	18.5925	18.7569	18.7595
14	19.0781	18.8312	19.0667	18.9019	19.0437	18.9152	18.8830
15	19.2117	19.1791	19.1920	19.2203	19.1567	19.3207	19.3096
16	19.7397	19.7592	19.7732	19.7231	19.7864	19.6585	19.6421
17	20.0453	19.8585	19.9699	19.8162	19.9213	19.8585	19.8007
18	20.1400	20.3112	20.2278	20.3539	20.2563	20.3426	20.3726
19	20.5006	20.5571	20.5544	20.4193	20.6025	20.4010	20.5801
20	20.8388	20.7738	20.7985	20.9092	20.7340	20.9533	20.7570
21	20.9949	21.1228	21.0618	21.1100	20.9733	21.0230	21.1956
22	21.4400	21.3062	21.3858	21.3206	21.4556	21.4300	21.2393
23	21.7341	21.7283	21.7123	21.5806	21.6312	21.5857	21.5180
24	21.7513	21.7482	21.7820	21.8952	21.8479	21.9129	21.9588
25	22.2290	22.2258	22.0692	22.1438	22.0942	22.2088	22.0964
26	22.2578	22.2527	22.4228	22.3350	22.3855	22.2915	22.3842
27	22.6882	22.6604	22.6099	22.6480	22.5001	22.5762	22.6563
28	22.7606	22.7806	22.8444	22.7936	22.9397	22.8829	22.7878
29	23.1487	23.1330	23.1023	23.1535	23.1422	23.1202	23.0904
30	23.2273	23.2357	23.2783	23.2158	23.2283	23.2658	23.2805
31	23.4658	23.4854	23.4980	23.4196	23.5960	23.6339	23.4294
32	23.8039	23.7781	23.7762	23.8423	23.6701	23.6466	23.8348
33	23.9973	24.0435	23.9706	23.9845	24.0055	23.9520	23.8593
34	24.1413	24.0892	24.1702	24.1478	24.1277	24.1931	24.2720
35	24.3597	24.4737	24.4272	24.4861	24.3148	24.3831	24.4410
36	24.6194	24.5008	24.5542	24.4881	24.6579	24.6021	24.5345
37	24.7825	24.7666	24.7845	24.7390	24.8295	24.8552	24.8831
38	25.0145	25.0248	25.0128	25.0513	24.9625	24.9468	24.9100
39	25.1304	25.1705	25.2924	25.1722	25.2928	25.2141	25.2447
40	25.4615	25.4173	25.3004	25.4149	25.2952	25.3819	25.3440
41	25.6829	25.6169	25.6321	25.5968	25.5392	25.5409	25.5130
42	25.6863	25.7476	25.7351	25.7666	25.8231	25.8285	25.8497
43	25.9377	25.9233	25.9191	25.9739	26.0029	25.9570	25.9653
44	26.1874	26.1977	26.2029	26.1466	26.1174	26.1686	26.1555
45	26.4264	26.3803	26.3597	26.3871	26.4099	26.4184	26.2639
46	26.4393	26.4816	26.5017	26.4734	26.4501	26.4460	26.5953
47	26.7116	26.6949	26.6556	26.7046	26.6541	26.6584	26.7175
48	26.8769	26.8903	26.9273	26.8790	26.9280	26.9270	26.8662
49	26.9891	26.9971	27.1137	27.0018	26.9852	26.9839	27.1297
50	27.3064	27.2957	27.1772	27.2893	27.3045	27.3078	27.1624
51	27.4150	27.4665	27.4273	27.4722	27.4582	27.4318	27.4055
52	27.5752	27.5217	27.5562	27.5139	27.5267	27.5537	27.5799
53	27.6795	27.6879	27.7920	27.7267	27.6682	27.7439	27.7088
54	27.9900	27.9798	27.8709	27.9393	27.9954	27.9204	27.9564
55	28.0524	28.0161	28.0165	28.0130	28.1150	28.0487	28.1333

Appendix II: Triangular (10, 30, 40)
100 generated values as per 25 runs in column A to Y

	S	T	U	V	W	X	Y
56	28.2852	28.3194	28.3119	28.3201	28.2175	28.2817	28.2002
57	28.4601	28.4838	28.4079	28.4346	28.3890	28.3445	28.3860
58	28.5337	28.5088	28.5759	28.5551	28.5986	28.6398	28.6024
59	28.8008	28.7830	28.6693	28.7207	28.7126	28.7157	28.6983
60	28.8371	28.8539	28.9570	28.9130	28.9191	28.9122	28.9341
61	28.9938	29.1006	29.0044	29.0482	29.0648	28.9962	28.9846
62	29.2764	29.1700	29.2545	29.2191	29.2004	29.2634	29.2807
63	29.3292	29.2940	29.3592	29.3099	29.4024	29.4187	29.3474
64	29.5642	29.5987	29.5223	29.5800	29.4859	29.4636	29.5415
65	29.6265	29.7204	29.6869	29.7169	29.6766	29.7361	29.6000
66	29.8801	29.7870	29.8070	29.7872	29.8247	29.7584	29.9013
67	29.9488	29.9810	29.9471	29.9518	29.9468	29.9841	29.9462
68	30.1636	30.1314	30.1512	30.1575	30.1598	30.1139	30.1613
69	30.2690	30.3569	30.3542	30.3481	30.3190	30.3163	30.2152
70	30.4567	30.3686	30.3556	30.3737	30.4000	30.3936	30.5067
71	30.5320	30.6229	30.6598	30.5671	30.5910	30.6407	30.5703
72	30.8284	30.7366	30.6825	30.7895	30.7619	30.7013	30.7840
73	30.9570	30.9383	30.9932	30.9655	30.9670	30.9979	30.9064
74	31.0569	31.0774	31.0037	31.0457	31.0408	30.9983	31.1034
75	31.3073	31.2172	31.2122	31.3411	31.2224	31.2983	31.2510
76	31.3868	31.4806	31.4660	31.3505	31.4673	31.3773	31.4390
77	31.6915	31.6228	31.5684	31.5794	31.6271	31.5368	31.6435
78	31.7106	31.7823	31.8169	31.8222	31.7697	31.8484	31.7540
79	32.0137	32.0098	32.0214	31.9224	31.9808	32.0468	31.9331
80	32.1284	32.1351	32.1009	32.2201	32.1559	32.0739	32.2059
81	32.3205	32.3287	32.2927	32.4357	32.4219	32.2671	32.3244
82	32.5996	32.5944	32.6073	32.4800	32.4898	32.6324	32.5903
83	32.8424	32.8272	32.7185	32.7232	32.7873	32.7475	32.8004
84	32.8930	32.9121	32.9977	33.0135	32.9428	32.9656	32.9303
85	33.1172	33.2409	33.2883	33.1666	33.1619	33.1552	33.1324
86	33.4909	33.3668	33.2915	33.4376	33.4377	33.4250	33.4695
87	33.6831	33.7539	33.5970	33.5355	33.5471	33.5713	33.5530
88	33.8484	33.7814	33.9130	34.0026	33.9849	33.9374	33.9794
89	34.2211	34.1401	34.0019	34.1741	34.1998	34.0302	34.1592
90	34.3119	34.4008	34.5164	34.3590	34.3272	34.4823	34.3698
91	34.6242	34.6955	34.7353	34.7208	34.6250	34.6916	34.7653
92	35.0140	34.9444	34.8690	34.9112	35.0064	34.9106	34.8603
93	35.4248	35.3743	35.2164	35.3199	35.1186	35.1960	35.1849
94	35.4347	35.4929	35.6230	35.5420	35.7567	35.6412	35.6822
95	36.1299	36.0105	35.8898	35.9276	35.7950	36.0022	35.8009
96	36.1507	36.2827	36.3748	36.3653	36.5121	36.2465	36.5055
97	36.9168	36.8300	36.7989	36.5744	36.6655	36.7042	36.5487
98	37.1153	37.2209	37.2104	37.5314	37.4015	37.3171	37.5587
99	38.1763	38.1297	37.7262	38.1167	37.9673	37.7471	38.1191
100	38.3993	38.4674	39.0805	38.4704	38.6635	39.0194	38.4563
101							
102							
103	mean						
104	26.6667	26.6667	26.6667	26.6667	26.6667	26.6667	26.6667
105	std.dev						
106	6.2361	6.2361	6.2361	6.2361	6.2361	6.2361	6.2361
107							

Appendix III

100 generated values as per 25 runs of N(100,36)

Appendix III: Normal (100, 36)
100 generated values as per 25 runs in column A to Y

	A	B	C	D	E	F	G	H	I
1	85.8737	84.5482	84.9738	85.3771	81.8872	84.0757	83.5384	82.6717	84.8797
2	86.8249	86.9071	86.5671	86.3486	87.6167	87.0078	87.2558	87.4446	86.6428
3	88.0376	88.0827	88.5783	88.0953	87.8767	88.1347	88.0651	87.7157	88.5964
4	89.7061	89.1616	88.6065	89.0427	89.5062	89.0660	89.2293	89.4940	88.6124
5	90.0729	89.6271	89.8879	89.4448	90.0690	89.4792	89.6784	89.9464	89.5921
6	90.6179	90.4965	90.1544	90.5393	90.3328	90.5629	90.5097	90.3308	90.4276
7	91.2327	90.6638	90.8492	90.6964	90.9085	91.0230	90.9905	90.8011	90.8939
8	91.4433	91.5040	91.2398	91.3882	91.4953	91.1428	91.2720	91.4873	91.2154
9	92.0531	91.7836	91.6319	91.4973	91.8331	91.8713	91.7673	91.6591	91.6893
10	92.1960	92.0499	92.0929	92.2234	92.1968	91.9286	92.1167	92.2596	92.0554
11	92.5762	92.3988	92.3223	92.3258	92.4847	92.5830	92.5839	92.5964	92.2648
12	92.9942	92.8062	92.7892	92.7972	92.9075	92.5987	92.6796	92.7135	92.8541
13	93.3650	93.0791	93.1007	92.9910	93.0294	93.1518	93.0965	93.0095	92.8859
14	93.3800	93.3375	93.2359	93.3507	93.5546	93.2408	93.3679	93.4928	93.4495
15	93.8469	93.7211	93.5127	93.5965	93.7404	93.5013	93.5187	93.6519	93.6355
16	93.9446	93.7823	93.9094	93.8390	93.9293	93.9704	94.0199	93.9382	93.8037
17	94.2995	94.1085	94.1408	94.1756	94.1672	94.2265	94.1803	94.2596	93.9840
18	94.4490	94.3832	94.2854	94.2592	94.4826	94.2497	94.3570	94.3221	94.4452
19	94.7982	94.6162	94.5884	94.4672	94.5629	94.6968	94.5967	94.5490	94.6537
20	94.8388	94.7962	94.7630	94.8859	94.9949	94.7019	94.8564	94.9432	94.7083
21	95.2093	95.0068	94.9367	95.0493	95.1630	95.0809	95.0511	95.0802	95.0663
22	95.2595	95.2657	95.2791	95.1759	95.2549	95.1821	95.2624	95.2749	95.1618
23	95.5151	95.3552	95.5116	95.4589	95.5094	95.4891	95.4398	95.5361	95.3785
24	95.7388	95.7294	95.5280	95.5854	95.7145	95.5913	95.6858	95.6316	95.6665
25	95.8401	95.7617	95.8752	95.7666	95.8318	95.7550	95.7671	95.9294	95.8901
26	96.1614	96.0994	95.9452	96.0558	96.1589	96.1008	96.1305	96.0114	95.9375
27	96.3531	96.1468	96.2219	96.1880	96.3390	96.2250	96.1463	96.1654	96.2287
28	96.3693	96.4582	96.3466	96.3834	96.3913	96.3796	96.4936	96.5133	96.3462
29	96.6725	96.5102	96.4785	96.6066	96.6904	96.5225	96.5464	96.6896	96.5895
30	96.7414	96.8112	96.8089	96.6859	96.7496	96.7999	96.8095	96.7061	96.7058
31	96.9917	96.9674	96.9017	96.9664	97.0450	97.0025	96.9228	96.8876	96.9834
32	97.0919	97.0493	97.0856	97.0235	97.0822	97.0176	97.1257	97.1971	97.0096
33	97.2709	97.2225	97.1816	97.3161	97.3054	97.3417	97.2472	97.2554	97.1954
34	97.4632	97.4672	97.4835	97.3518	97.4893	97.3550	97.4737	97.5021	97.4743
35	97.5298	97.5335	97.6088	97.5749	97.6471	97.6301	97.5905	97.6192	97.5639
36	97.8389	97.8138	97.7202	97.7543	97.7999	97.7268	97.7872	97.7947	97.7683
37	97.8675	97.8429	97.9534	97.8690	97.9225	97.9845	97.9956	97.9394	97.8494
38	98.1241	98.1486	98.0245	98.1080	98.1626	98.0193	98.0256	98.1167	98.1305
39	98.1749	98.2646	98.2216	98.2033	98.2998	98.1634	98.2614	98.2900	98.1808
40	98.4284	98.3611	98.3935	98.4107	98.4134	98.4747	98.3913	98.3975	98.4362
41	98.5299	98.6007	98.4666	98.5851	98.5927	98.5992	98.5738	98.6095	98.5813
42	98.6767	98.6489	98.7760	98.6570	98.7387	98.6663	98.7015	98.7000	98.6640
43	98.8130	98.8590	98.8598	98.9095	98.9073	98.8648	98.8530	98.9495	98.9033
44	98.9891	99.0071	99.0046	98.9528	99.0354	99.0195	99.0375	98.9747	98.9624
45	99.0647	99.2083	99.1804	99.1117	99.1886	99.2426	99.2064	99.2495	99.1894
46	99.3276	99.2690	99.2993	99.3649	99.3597	99.2553	99.2941	99.2838	99.2910
47	99.3836	99.5398	99.5248	99.4551	99.4640	99.5100	99.5496	99.4352	99.4365
48	99.5954	99.5447	99.5664	99.6326	99.6861	99.5975	99.5568	99.7033	99.6547
49	99.7296	99.7729	99.7460	99.7440	99.7697	99.7406	99.7417	99.7540	99.7795
50	99.8338	99.9165	99.9544	99.9523	99.9801	99.9740	99.9683	99.9875	99.9204
51	100.0511	100.0057	100.0445	100.0256	100.1088	100.1260	100.0308	100.1654	100.0649
52	100.0951	100.2871	100.2634	100.2777	100.2390	100.1943	100.2815	100.1776	100.2418
53	100.2918	100.3920	100.3995	100.4027	100.4455	100.3930	100.3655	100.3563	100.3121
54	100.4387	100.5056	100.5175	100.5089	100.5018	100.5346	100.5503	100.5899	100.6034
55	100.5392	100.6504	100.6670	100.6086	100.6778	100.6183	100.6427	100.7679	100.7309

Appendix III: Normal (100, 36)
100 generated values as per 25 runs in column A to Y

	A	B	C	D	E	F	G	H	I
56	100.7781	100.8545	100.8616	100.9142	100.8716	100.9193	100.8793	100.7834	100.7952
57	100.9383	100.9474	101.0016	101.0308	101.0088	100.9953	101.0546	100.9888	100.9728
58	100.9690	101.1688	101.1422	101.1063	101.1461	101.1554	101.0768	101.1719	101.1683
59	101.1577	101.3171	101.3256	101.3035	101.2959	101.3536	101.2403	101.3293	101.2460
60	101.3455	101.4154	101.4391	101.4540	101.4703	101.4160	101.5071	101.4460	101.5158
61	101.4504	101.5844	101.6421	101.6691	101.6300	101.6013	101.6392	101.6845	101.6196
62	101.6556	101.7722	101.7508	101.7157	101.7545	101.7949	101.7302	101.7125	101.7694
63	101.8232	101.8966	101.8794	101.9982	101.8606	102.0049	101.9722	101.8761	101.9651
64	101.8939	102.0931	102.1519	102.0234	102.1523	102.0264	102.0288	102.1530	102.0606
65	102.0473	102.1880	102.2885	102.2822	102.2268	102.2675	102.2706	102.2394	102.3056
66	102.2931	102.4465	102.3906	102.3876	102.4242	102.4108	102.3735	102.4315	102.3682
67	102.3988	102.6250	102.6613	102.6578	102.5139	102.5533	102.5581	102.5320	102.5872
68	102.5766	102.6663	102.6801	102.6735	102.7898	102.7860	102.7428	102.7954	102.7487
69	102.6876	102.8385	102.9708	102.8649	102.9696	102.9873	102.8711	102.9313	102.9205
70	102.9397	103.1284	103.0494	103.1460	103.0008	103.0277	103.1029	103.0671	103.0936
71	103.1271	103.2244	103.3491	103.2253	103.3109	103.2972	103.2716	103.3416	103.2995
72	103.1685	103.4351	103.3691	103.4829	103.3467	103.4140	103.3935	103.3475	103.4117
73	103.4253	103.6163	103.5649	103.6876	103.6745	103.7010	103.5914	103.5790	103.6137
74	103.5624	103.7588	103.8769	103.7395	103.6928	103.7288	103.7889	103.8254	103.8186
75	103.7387	103.9738	103.9613	104.0805	103.9900	104.0402	103.9804	103.9386	103.9282
76	103.9679	104.1447	104.2283	104.0940	104.1142	104.1359	104.1413	104.2072	104.2536
77	104.1109	104.3931	104.4297	104.4259	104.2550	104.3190	104.4118	104.2869	104.3365
78	104.3450	104.5004	104.5392	104.5290	104.6216	104.6384	104.4831	104.6334	104.6250
79	104.5922	104.6726	104.7386	104.6919	104.7640	104.6814	104.7964	104.7284	104.8009
80	104.6486	105.0392	105.0531	105.0868	104.9162	105.0960	104.9113	105.0025	104.9783
81	104.9481	105.1360	105.3171	105.1736	105.1239	105.1570	105.2342	105.2288	105.1648
82	105.1257	105.4357	105.3385	105.4695	105.4122	105.4832	105.3328	105.3587	105.4833
83	105.4644	105.5117	105.7790	105.6024	105.7048	105.7425	105.5677	105.7503	105.6107
84	105.4959	105.9845	105.8021	105.9671	105.7390	105.8165	105.9209	105.7526	105.9631
85	105.8396	106.1836	106.2506	106.0345	106.0143	106.2173	106.0069	106.1623	106.0809
86	106.0807	106.2955	106.3289	106.5377	106.4191	106.3372	106.4753	106.3299	106.4924
87	106.3013	106.7589	106.8077	106.5731	106.6684	106.8132	106.5550	106.7677	106.5994
88	106.6700	106.8035	106.8632	107.0910	106.8334	106.8289	107.0090	106.8038	107.0671
89	107.0245	107.1648	107.1879	107.3536	107.3060	107.3075	107.2884	107.2111	107.2827
90	107.1059	107.6140	107.7116	107.5128	107.3907	107.5487	107.4716	107.5695	107.5922
91	107.4719	107.7939	108.1086	108.0062	107.8054	107.8692	107.9189	107.8938	108.0031
92	108.0038	108.3690	108.1631	108.2463	108.2668	108.3802	108.2176	108.2595	108.2548
93	108.2068	108.8645	108.5725	108.9022	108.4504	108.9293	108.8241	108.5690	108.9333
94	108.8503	108.9066	109.3766	108.9839	109.2630	108.9391	108.9315	109.2249	108.9576
95	109.0632	109.4579	109.8602	109.4881	109.6579	109.7704	109.7049	109.3727	109.5604
96	109.9961	110.4109	110.1048	110.5177	110.0433	110.1528	110.0923	110.5137	110.4303
97	110.3574	110.8451	111.1793	111.3195	110.7970	111.1123	110.5832	111.1328	110.9160
98	111.4880	111.8864	111.6620	111.4744	111.7870	111.6797	112.2563	111.5076	111.9671
99	111.8799	112.8535	113.9667	113.4289	113.0444	113.8775	113.0026	113.6261	113.7300
100	122.4335	116.1465	114.2686	114.9839	115.2307	114.2948	115.6029	114.4045	114.5195
101									
102									
103	mean								
104	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
105	std.dev								
106	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000
107									

Appendix III: Normal (100, 36)
100 generated values as per 25 runs in column A to Y

	J	K	L	M	N	O	P	Q	R
1	85.4337	81.5679	85.0991	85.6860	84.9158	83.6198	85.5601	85.6243	85.7513
2	86.5008	87.8251	86.5132	86.2020	86.6244	87.2158	86.2781	86.2850	86.2455
3	87.6789	88.4308	88.3870	88.1973	87.8477	88.5170	88.2718	88.0094	87.7555
4	89.4611	89.3834	88.7959	89.0301	89.2181	88.8458	88.9521	89.1958	89.4032
5	89.6423	89.8193	89.4649	89.5669	89.7638	89.7048	89.5640	89.5505	89.9237
6	90.5229	90.7765	90.5158	90.5006	90.2873	90.4661	90.4867	90.5393	90.2718
7	90.9122	91.3198	90.7371	90.9828	90.6664	90.7160	90.8809	90.7769	90.7311
8	91.3248	91.3278	91.3455	91.1812	91.4179	91.4923	91.2607	91.3929	91.4700
9	91.9033	91.8620	91.5407	91.8863	91.8250	91.6595	91.5865	91.7894	91.5444
10	91.9650	92.3595	92.1793	91.9093	91.9297	92.1951	92.1758	92.0272	92.2821
11	92.4467	92.7555	92.4785	92.4621	92.4477	92.3887	92.4532	92.2769	92.3489
12	92.7872	92.8231	92.6477	92.7089	92.6892	92.8462	92.7052	92.8993	92.8665
13	93.0637	93.2953	93.0285	92.9626	93.0359	92.9797	92.9595	93.1532	93.0158
14	93.3738	93.4599	93.3099	93.4104	93.3141	93.4606	93.4027	93.2488	93.4090
15	93.6355	93.6666	93.5843	93.5323	93.5170	93.5856	93.6966	93.5600	93.7244
16	93.8839	94.1381	93.8458	93.9289	93.9201	93.9430	93.7615	93.9206	93.7883
17	94.1205	94.2194	94.1700	94.0878	94.0130	94.0824	94.1326	94.2273	94.1554
18	94.3865	94.5580	94.2604	94.3699	94.4218	94.4372	94.3183	94.2501	94.3443
19	94.6620	94.7778	94.5587	94.4948	94.6063	94.5286	94.6375	94.6445	94.6131
20	94.7632	94.9007	94.7950	94.8810	94.7588	94.9089	94.7360	94.7514	94.8031
21	94.9677	95.1332	94.9786	94.9975	95.1114	95.0424	95.0836	94.9719	95.1195
22	95.3120	95.3844	95.2418	95.2443	95.1209	95.2593	95.1537	95.2824	95.1566
23	95.4896	95.6205	95.5114	95.4630	95.4049	95.5371	95.5141	95.4244	95.3620
24	95.6050	95.6939	95.5308	95.5958	95.6449	95.5797	95.5399	95.6458	95.7221
25	95.7555	96.0227	95.8251	95.7867	95.7601	95.7681	95.7278	95.8352	95.7669
26	96.1087	96.0488	95.9969	96.0474	96.0690	96.1200	96.0999	96.0110	96.0915
27	96.2221	96.2769	96.1389	96.1188	96.1206	96.1760	96.1077	96.1563	96.2385
28	96.3855	96.5186	96.4298	96.4588	96.4560	96.4562	96.4659	96.4320	96.3633
29	96.5045	96.5904	96.5688	96.6304	96.5218	96.6491	96.6364	96.5365	96.6213
30	96.8151	96.9032	96.7223	96.6673	96.7766	96.7005	96.6586	96.7685	96.6948
31	96.8872	96.9845	96.8962	96.9233	96.9974	96.8699	96.8897	96.9705	96.8822
32	97.1242	97.1874	97.0926	97.0683	96.9999	97.1707	97.0996	97.0289	97.1240
33	97.2218	97.3088	97.2004	97.2900	97.1946	97.2926	97.2696	97.2888	97.2189
34	97.4614	97.5210	97.4663	97.3770	97.4794	97.4230	97.3963	97.3846	97.4596
35	97.5347	97.7106	97.5373	97.5934	97.5433	97.5603	97.5990	97.6588	97.5554
36	97.8039	97.7622	97.7918	97.7324	97.7930	97.8119	97.7266	97.6726	97.7792
37	97.9496	97.9702	97.9591	97.8718	97.9909	97.8971	97.8686	97.9438	97.9768
38	98.0321	98.1311	98.0195	98.0989	97.9945	98.1192	98.1026	98.0317	98.0010
39	98.1841	98.3030	98.1595	98.2899	98.2518	98.2795	98.2006	98.2021	98.2822
40	98.4278	98.4164	98.4554	98.3156	98.3704	98.3699	98.4057	98.4060	98.3268
41	98.5221	98.6073	98.5343	98.5201	98.5049	98.5201	98.4786	98.5195	98.6028
42	98.7117	98.7212	98.7094	98.7099	98.7446	98.7522	98.7531	98.7125	98.6283
43	98.8490	98.8572	98.8330	98.9002	98.8276	98.8340	98.8060	98.8508	98.8607
44	98.9994	99.0730	99.0313	98.9477	99.0425	99.0546	99.0443	98.9980	98.9850
45	99.2257	99.2025	99.1105	99.0861	99.1042	99.1643	99.1324	99.1711	99.1801
46	99.2319	99.3244	99.3689	99.3734	99.3806	99.3351	99.3310	99.2888	99.2750
47	99.5301	99.4513	99.3972	99.3990	99.4903	99.5016	99.4704	99.4132	99.5268
48	99.5325	99.6684	99.6934	99.6689	99.6057	99.6046	99.6021	99.6538	99.5338
49	99.8118	99.7229	99.8191	99.7720	99.7796	99.7948	99.7385	99.7606	99.7689
50	99.8536	99.9872	99.8807	99.9020	99.9251	99.9158	99.9406	99.9115	99.8948
51	100.1291	100.1392	100.1194	100.0297	100.1324	100.1432	100.0237	100.0374	100.1001
52	100.1374	100.1599	100.1876	100.2487	100.1791	100.1702	100.2605	100.2378	100.1650
53	100.3099	100.3888	100.4082	100.3039	100.3819	100.4248	100.3990	100.4179	100.4054
54	100.5595	100.5008	100.5078	100.5805	100.5383	100.4931	100.4917	100.4623	100.4629
55	100.7224	100.6629	100.6748	100.6400	100.7131	100.6450	100.7206	100.6929	100.7025

Appendix III: Normal (100, 36)
100 generated values as per 25 runs in column A to Y

	J	K	L	M	N	O	P	Q	R
56	100.7519	100.8195	100.8525	100.8526	100.8180	100.8798	100.7791	100.7944	100.7713
57	101.0271	101.0132	100.9249	100.9596	100.9276	100.9784	100.9071	101.0225	101.0343
58	101.0561	101.0656	101.2180	101.1451	101.2189	101.1570	101.2060	101.0759	101.0488
59	101.3122	101.3378	101.3785	101.3363	101.3150	101.2421	101.2293	101.2784	101.2938
60	101.3855	101.3429	101.3845	101.3858	101.4513	101.5098	101.5022	101.4367	101.4042
61	101.5476	101.5008	101.6024	101.5978	101.6372	101.5795	101.5628	101.5537	101.5936
62	101.7723	101.7899	101.7889	101.7493	101.7568	101.7957	101.7943	101.7857	101.7265
63	101.8774	101.8958	101.9185	101.9732	101.9588	101.8809	101.9162	101.8423	101.9288
64	102.0732	102.0120	102.1102	102.0078	102.0721	102.1271	102.0754	102.1306	102.0223
65	102.1807	102.1461	102.3003	102.1895	102.2020	102.2989	102.3153	102.2794	102.2505
66	102.4122	102.3915	102.3765	102.4378	102.4781	102.3520	102.3221	102.3367	102.3431
67	102.5231	102.4979	102.5168	102.5988	102.5548	102.6330	102.5522	102.6165	102.6121
68	102.7252	102.6814	102.8237	102.6865	102.7869	102.6753	102.7456	102.6573	102.6372
69	102.8473	102.9137	102.9397	102.9488	102.9252	102.8828	102.8454	102.8381	102.8383
70	103.0729	102.9230	103.0780	103.0116	103.0942	103.0999	103.1292	103.1112	103.0843
71	103.2447	103.1400	103.2873	103.2025	103.3376	103.2545	103.2690	103.1836	103.1682
72	103.3656	103.3746	103.4284	103.4536	103.3788	103.4205	103.3996	103.4591	103.4462
73	103.5558	103.4477	103.5613	103.5822	103.5372	103.6606	103.5671	103.6596	103.5968
74	103.7689	103.7673	103.8772	103.7908	103.9030	103.7294	103.8209	103.6974	103.7303
75	103.9269	103.8267	103.9627	103.9219	103.9687	103.9859	103.9831	103.9762	103.8742
76	104.1385	104.1135	104.2232	104.1964	104.2175	104.1473	104.1493	104.1246	104.1963
77	104.2511	104.2727	104.3043	104.2709	104.4617	104.3822	104.3808	104.3219	104.4049
78	104.5892	104.4228	104.6642	104.6259	104.5030	104.5257	104.5291	104.5554	104.4360
79	104.7433	104.6669	104.8370	104.8394	104.7736	104.8237	104.8344	104.7659	104.7385
80	104.9058	104.8235	104.9476	104.8691	105.0122	104.8973	104.8915	104.9247	104.9152
81	105.1794	105.1209	105.1193	105.1135	105.2063	105.2505	105.0868	105.2375	105.2287
82	105.3277	105.2096	105.5381	105.4620	105.4461	105.3311	105.5091	105.3136	105.2828
83	105.6142	105.5269	105.7411	105.5185	105.7617	105.5433	105.5744	105.6496	105.6226
84	105.8094	105.7016	105.8354	105.9815	105.8133	105.9633	105.9419	105.8215	105.8064
85	106.0122	105.9055	106.0612	106.1779	106.2168	106.1324	106.0365	105.9997	105.9999
86	106.4043	106.2960	106.5217	106.3071	106.3561	106.3599	106.4767	106.4710	106.4239
87	106.7336	106.5143	106.6966	106.6573	106.7656	106.6881	106.7844	106.6417	106.5021
88	106.7562	106.7410	106.9717	106.9158	106.8977	106.8867	106.8075	106.9065	107.0066
89	107.3357	107.1113	107.2716	107.3791	107.1789	107.3299	107.0935	107.2160	107.2262
90	107.3563	107.3216	107.6135	107.3992	107.7118	107.4493	107.7286	107.5408	107.4768
91	107.9834	107.8428	107.9725	107.7928	108.0733	108.0586	108.0362	107.9406	107.9114
92	108.0805	107.9315	108.2977	108.3831	108.1883	108.0950	108.1461	108.1904	108.1642
93	108.4514	108.3210	108.5782	108.5359	108.8860	108.8234	108.4877	108.5364	108.6651
94	109.2746	109.0765	109.3608	109.2842	109.0126	108.9567	109.3729	109.2490	109.0377
95	109.6455	109.1837	109.8554	109.7321	109.7849	109.7090	109.5900	109.8068	109.5897
96	110.0812	110.2443	110.1000	110.1027	110.1718	110.1160	110.2989	109.9863	110.1575
97	110.9574	110.8727	111.3169	111.0170	111.2857	110.8606	111.1899	111.1464	111.1018
98	111.6222	111.2622	111.4966	111.6892	111.5241	111.8718	111.5220	111.4958	111.4677
99	112.5483	112.1944	113.6731	112.7890	113.7406	113.0306	112.8951	112.6976	112.5588
100	117.3851	118.2146	114.6166	116.4948	114.5168	115.5955	116.1568	116.8225	117.3924
101									
102									
103	mean								
104	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
105	std.dev								
106	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000
107									

Appendix III: Normal (100, 36)

100 generated values as per 25 runs in column A to Y

	S	T	U	V	W	X	Y
1	84.3792	85.0299	83.9992	84.9460	85.1264	83.3359	85.1747
2	86.8874	86.5416	87.1494	86.6138	86.5347	87.3397	86.5007
3	88.1703	88.5165	88.0763	88.3242	88.4005	88.6882	87.9786
4	89.0170	88.6733	89.2302	88.8671	88.8186	88.7681	89.1412
5	89.4486	89.6779	90.0089	89.8989	89.5850	89.7487	89.6263
6	90.5654	90.3453	90.2298	90.1675	90.4507	90.4969	90.4109
7	91.0642	90.6378	91.0326	90.9693	90.6308	91.0181	90.6479
8	91.0807	91.4281	91.2362	91.1478	91.4594	91.2876	91.4394
9	91.6416	91.6610	91.7877	91.5379	91.6355	91.6634	91.5887
10	92.1253	92.0713	92.0994	92.1932	92.1177	92.2465	92.1538
11	92.4987	92.4678	92.3087	92.3574	92.3471	92.4916	92.3589
12	92.6627	92.6575	92.9352	92.7726	92.7915	92.8017	92.7756
13	92.9657	93.1591	93.0307	93.1053	92.9871	93.1829	93.1258
14	93.4007	93.1826	93.4281	93.2457	93.3667	93.3150	93.2297
15	93.5116	93.4733	93.5316	93.5110	93.4607	93.6495	93.5847
16	93.9448	93.9497	94.0051	93.9236	93.9771	93.9244	93.8575
17	94.1921	94.0304	94.1632	93.9992	94.0864	94.0856	93.9865
18	94.2682	94.3950	94.3692	94.4320	94.3557	94.4720	94.4469
19	94.5566	94.5913	94.6282	94.4842	94.6318	94.5182	94.6122
20	94.8248	94.7633	94.8204	94.8729	94.7360	94.9527	94.7525
21	94.9479	95.0389	95.0267	95.0312	94.9250	95.0073	95.0984
22	95.2974	95.1830	95.2793	95.1965	95.3035	95.3242	95.1327
23	95.5271	95.5133	95.5328	95.4000	95.4407	95.4450	95.3510
24	95.5405	95.5289	95.5867	95.6455	95.6096	95.6981	95.6950
25	95.9123	95.9009	95.8087	95.8390	95.8013	95.9264	95.8020
26	95.9347	95.9218	96.0815	95.9876	96.0275	95.9901	96.0257
27	96.2688	96.2386	96.2257	96.2305	96.1164	96.2094	96.2369
28	96.3249	96.3320	96.4064	96.3436	96.4574	96.4454	96.3389
29	96.6261	96.6058	96.6052	96.6230	96.6146	96.6281	96.5738
30	96.6872	96.6857	96.7409	96.6715	96.6814	96.7403	96.7216
31	96.8726	96.8801	96.9106	96.8300	96.9674	97.0244	96.8374
32	97.1360	97.1084	97.1258	97.1595	97.0251	97.0342	97.1534
33	97.2872	97.3161	97.2767	97.2706	97.2870	97.2706	97.1725
34	97.4000	97.3519	97.4319	97.3985	97.3826	97.4580	97.4959
35	97.5715	97.6544	97.6325	97.6647	97.5295	97.6060	97.6289
36	97.7763	97.6758	97.7319	97.6663	97.7999	97.7773	97.7026
37	97.9054	97.8863	97.9129	97.8647	97.9359	97.9763	97.9789
38	98.0900	98.0919	98.0933	98.1132	98.0417	98.0486	98.0003
39	98.1825	98.2084	98.3156	98.2099	98.3059	98.2604	98.2679
40	98.4486	98.4070	98.3219	98.4050	98.3078	98.3942	98.3476
41	98.6281	98.5687	98.5880	98.5522	98.5045	98.5215	98.4840
42	98.6308	98.6751	98.6711	98.6903	98.7354	98.7535	98.7580
43	98.8362	98.8189	98.8205	98.8601	98.8827	98.8579	98.8527
44	99.0421	99.0453	99.0528	99.0025	98.9771	99.0310	99.0096
45	99.2409	99.1974	99.1823	99.2025	99.2200	99.2371	99.0995
46	99.2517	99.2823	99.3003	99.2747	99.2537	99.2600	99.3769
47	99.4811	99.4622	99.4290	99.4696	99.4253	99.4372	99.4803
48	99.6217	99.6286	99.6586	99.6181	99.6583	99.6640	99.6069
49	99.7179	99.7203	99.8180	99.7233	99.7074	99.7124	99.8335
50	99.9926	99.9793	99.8726	99.9726	99.9838	99.9907	99.8618
51	100.0876	100.1291	100.0893	100.1329	100.1184	100.0983	100.0737
52	100.2293	100.1779	100.2024	100.1698	100.1790	100.2053	100.2280
53	100.3222	100.3260	100.4118	100.3594	100.3049	100.3737	100.3431
54	100.6028	100.5901	100.4827	100.5516	100.6005	100.5319	100.5668
55	100.6599	100.6233	100.6143	100.6189	100.7102	100.6480	100.7291

Appendix III: Normal (100. 36)
100 generated values as per 25 runs in column A to Y

	S	T	U	V	W	X	Y
56	100.8753	100.9042	100.8857	100.9031	100.8050	100.8619	100.7910
57	101.0396	101.0592	100.9752	101.0108	100.9652	100.9202	100.9646
58	101.1095	101.0830	101.1335	101.1253	101.1641	101.1981	101.1701
59	101.3666	101.3468	101.2225	101.2843	101.2737	101.2708	101.2623
60	101.4021	101.4161	101.5012	101.4720	101.4751	101.4609	101.4924
61	101.5563	101.6606	101.5479	101.6060	101.6195	101.5432	101.5423
62	101.8403	101.7304	101.7975	101.7778	101.7558	101.8097	101.8400
63	101.8943	101.8564	101.9039	101.8703	101.9623	101.9679	101.9084
64	102.1382	102.1731	102.0721	102.1508	102.0489	102.0141	102.1099
65	102.2039	102.3025	102.2448	102.2962	102.2498	102.2996	102.1715
66	102.4765	102.3742	102.3730	102.3717	102.4089	102.3234	102.4955
67	102.5518	102.5863	102.5249	102.5512	102.5423	102.5677	102.5448
68	102.7895	102.7530	102.7491	102.7790	102.7778	102.7103	102.7828
69	102.9058	103.0022	102.9716	102.9895	102.9536	102.9320	102.8425
70	103.1126	103.0150	102.9732	103.0176	103.0428	103.0165	103.1638
71	103.1955	103.2952	103.3055	103.2306	103.2529	103.2861	103.2337
72	103.5212	103.4202	103.3302	103.4750	103.4405	103.3521	103.4686
73	103.6623	103.6418	103.6689	103.6682	103.6656	103.6751	103.6029
74	103.7719	103.7946	103.6804	103.7562	103.7466	103.6756	103.8191
75	104.0468	103.9483	103.9076	104.0806	103.9458	104.0021	103.9812
76	104.1341	104.2380	104.1845	104.0909	104.2148	104.0883	104.1878
77	104.4695	104.3947	104.2964	104.3428	104.3906	104.2622	104.4129
78	104.4905	104.5708	104.5685	104.6107	104.5477	104.6032	104.5349
79	104.8257	104.8228	104.7933	104.7216	104.7812	104.8211	104.7329
80	104.9530	104.9620	104.8810	105.0522	104.9756	104.8509	105.0360
81	105.1670	105.1780	105.0932	105.2932	105.2724	105.0643	105.1682
82	105.4801	105.4763	105.4436	105.3429	105.3485	105.4707	105.4665
83	105.7549	105.7400	105.5682	105.6170	105.6841	105.5998	105.7041
84	105.8124	105.8366	105.8838	105.9474	105.8609	105.8459	105.8519
85	106.0688	106.2143	106.2161	106.1233	106.1120	106.0617	106.0835
86	106.5024	106.3605	106.2198	106.4378	106.4319	106.3721	106.4747
87	106.7288	106.8162	106.5745	106.5524	106.5600	106.5423	106.5725
88	106.9254	106.8489	106.9480	107.1088	107.0810	106.9746	107.0804
89	107.3769	107.2817	107.0544	107.3174	107.3424	107.0858	107.2988
90	107.4886	107.6032	107.6841	107.5452	107.4992	107.6384	107.5582
91	107.8794	107.9747	107.9600	108.0007	107.8721	107.9011	108.0574
92	108.3824	108.2961	108.1313	108.2464	108.3640	108.1813	108.1801
93	108.9345	108.8706	108.5869	108.7899	108.5122	108.5556	108.6075
94	108.9482	109.0340	109.1428	109.0955	109.3916	109.1637	109.2927
95	109.9489	109.7762	109.5232	109.6468	109.4467	109.6826	109.4625
96	109.9803	110.1888	110.2535	110.3104	110.5361	110.0496	110.5333
97	111.2209	111.0779	110.9428	110.6446	110.7865	110.7782	110.6030
98	111.5736	111.7747	111.6703	112.3691	112.1038	111.8641	112.4254
99	113.8110	113.7023	112.6932	113.6638	113.3017	112.7312	113.6717
100	114.3950	114.5916	116.6071	114.5942	115.1659	116.3505	114.5575
101							
102							
103	mean						
104	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
105	std.dev						
106	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000	6.0000
107							

Appendix IV

100 generated values as per 25 runs of B(12,15,30)

Appendix IV: Beta (12, 15, 30)
100 generated values as per 25 runs in column A to Y

	A	B	C	D	E	F	G	H	I
1	12.6418	12.4595	12.5260	12.4663	12.5661	12.5650	12.5910	12.1845	12.5500
2	12.7258	12.7377	12.7636	12.7410	12.7013	12.6856	12.6286	12.8184	12.7406
3	12.9772	12.9486	12.8492	12.8969	12.9221	12.9151	12.8898	12.8529	12.9476
4	13.0446	12.9626	13.1070	13.0216	13.0116	13.0050	13.0022	13.0658	13.0111
5	13.2144	13.0953	13.1700	13.1397	13.1673	13.1709	13.0971	13.1729	13.1725
6	13.2342	13.2526	13.2376	13.2152	13.2013	13.1854	13.2337	13.1933	13.2188
7	13.3340	13.2927	13.3683	13.3203	13.3523	13.2879	13.3091	13.2955	13.3578
8	13.4581	13.4075	13.3862	13.3859	13.3663	13.4196	13.3748	13.4216	13.3816
9	13.5004	13.4748	13.4809	13.4895	13.4647	13.4453	13.4718	13.5079	13.5057
10	13.5897	13.5306	13.5745	13.5214	13.5576	13.5668	13.5185	13.5136	13.5357
11	13.6526	13.5912	13.6297	13.5964	13.6435	13.6108	13.6029	13.5825	13.6114
12	13.7079	13.6889	13.6982	13.6890	13.6542	13.6771	13.6651	13.7111	13.7022
13	13.7642	13.7244	13.7834	13.7274	13.7630	13.7452	13.6980	13.7613	13.7237
14	13.8426	13.8099	13.7952	13.8116	13.7871	13.7961	13.8215	13.7894	13.8398
15	13.8743	13.8335	13.8785	13.8563	13.8765	13.8767	13.8758	13.8726	13.8652
16	13.9645	13.9387	13.9330	13.9201	13.9086	13.9004	13.8842	13.9137	13.9351
17	14.0086	13.9443	13.9896	13.9770	14.0001	13.9881	13.9758	13.9940	13.9691
18	14.0531	14.0541	14.0468	14.0273	14.0120	14.0167	14.0132	14.0198	14.0558
19	14.1169	14.1061	14.0817	14.0737	14.0924	14.0968	14.0979	14.0989	14.0780
20	14.1560	14.1107	14.1683	14.1466	14.1351	14.1240	14.1084	14.1309	14.1623
21	14.1959	14.1997	14.2065	14.1892	14.1672	14.2098	14.1599	14.2037	14.1749
22	14.2821	14.2271	14.2506	14.2408	14.2691	14.2205	14.2572	14.2356	14.2732
23	14.2922	14.2866	14.2911	14.2809	14.3012	14.3089	14.2786	14.2828	14.2978
24	14.3852	14.3450	14.3676	14.3533	14.3393	14.3262	14.3440	14.3604	14.3531
25	14.4187	14.3888	14.4041	14.3770	14.4183	14.3715	14.4038	14.3859	14.3875
26	14.4539	14.4421	14.4515	14.4564	14.4201	14.4620	14.4190	14.4564	14.4606
27	14.5249	14.4875	14.5126	14.4695	14.4763	14.4958	14.4666	14.4955	14.5027
28	14.5403	14.5401	14.5373	14.5599	14.5574	14.5342	14.5531	14.5431	14.5403
29	14.6097	14.6034	14.5751	14.5789	14.5685	14.6057	14.5739	14.5861	14.5908
30	14.6434	14.6172	14.6648	14.6435	14.6576	14.6166	14.6401	14.6449	14.6428
31	14.6934	14.6781	14.6936	14.6998	14.6821	14.6793	14.6621	14.6646	14.6931
32	14.7468	14.7341	14.7353	14.7143	14.7346	14.7341	14.7439	14.7572	14.7303
33	14.7750	14.7869	14.7821	14.7834	14.7897	14.7952	14.7539	14.7821	14.7961
34	14.8501	14.8147	14.8336	14.8196	14.8155	14.8073	14.8426	14.8290	14.8147
35	14.8816	14.8659	14.8785	14.8843	14.8769	14.8859	14.8929	14.8681	14.8616
36	14.9280	14.9246	14.9234	14.9073	14.9162	14.9052	14.8933	14.9314	14.9361
37	14.9954	14.9605	14.9592	14.9742	14.9485	14.9848	14.9548	14.9827	14.9557
38	14.9979	15.0182	15.0283	15.0051	15.0317	14.9938	15.0201	15.0045	15.0282
39	15.0779	15.0475	15.0701	15.0773	15.0687	15.0772	15.0586	15.0708	15.0823
40	15.0991	15.1194	15.1029	15.0899	15.0988	15.0893	15.1052	15.1041	15.0879
41	15.1521	15.1413	15.1483	15.1610	15.1390	15.1426	15.1383	15.1626	15.1378
42	15.2091	15.2141	15.2106	15.1945	15.2160	15.2119	15.2148	15.2004	15.2190
43	15.2490	15.2281	15.2558	15.2560	15.2246	15.2548	15.2344	15.2487	15.2465
44	15.2970	15.3167	15.2898	15.2885	15.3190	15.2886	15.3088	15.3031	15.2976
45	15.3450	15.3576	15.3589	15.3320	15.3547	15.3437	15.3226	15.3503	15.3652
46	15.3871	15.3776	15.3747	15.4028	15.3783	15.3898	15.4119	15.3917	15.3675
47	15.4444	15.4551	15.4297	15.4382	15.4275	15.4528	15.4235	15.4287	15.4302
48	15.4753	15.4723	15.4934	15.4884	15.4967	15.4725	15.5039	15.5050	15.4927
49	15.5091	15.5284	15.5218	15.5571	15.5125	15.5337	15.5422	15.5287	15.5159
50	15.6004	15.5932	15.5927	15.5633	15.6050	15.5853	15.5800	15.5986	15.5992
51	15.6337	15.6304	15.6113	15.6395	15.6383	15.6254	15.6572	15.6367	15.6058
52	15.6679	15.6881	15.6974	15.6775	15.6749	15.6901	15.6627	15.6870	15.7042
53	15.7431	15.7584	15.7431	15.7146	15.7286	15.7428	15.7198	15.7262	15.7496
54	15.7528	15.7590	15.7619	15.8018	15.7828	15.7712	15.8005	15.7963	15.7573
55	15.8144	15.8452	15.8205	15.8527	15.8404	15.8343	15.8175	15.8358	15.8049

Appendix IV: Beta (12, 15, 30)
100 generated values as per 25 runs in column A to Y

	A	B	C	D	E	F	G	H	I
56	15.8805	15.8761	15.8855	15.8666	15.8736	15.8831	15.9068	15.8898	15.9037
57	15.9035	15.9372	15.9329	15.9129	15.9402	15.9388	15.9614	15.9327	15.9450
58	15.9936	15.9905	15.9764	16.0132	15.9789	15.9843	15.9692	15.9989	15.9673
59	16.0219	16.0610	16.0201	16.0577	16.0575	16.0390	16.0551	16.0647	16.0211
60	16.0817	16.0788	16.0985	16.0793	16.0729	16.0959	16.0893	16.0783	16.1017
61	16.1448	16.1413	16.1329	16.1408	16.1237	16.1408	16.1373	16.1505	16.1633
62	16.1696	16.2146	16.1988	16.2119	16.2222	16.2096	16.2244	16.2079	16.1733
63	16.2310	16.2420	16.2694	16.2380	16.2281	16.2711	16.2728	16.2455	16.2371
64	16.3007	16.3368	16.2816	16.3374	16.3394	16.3013	16.3117	16.3353	16.3199
65	16.3446	16.3986	16.3502	16.3863	16.3421	16.3774	16.4016	16.3826	16.3820
66	16.4116	16.4096	16.4276	16.4180	16.4532	16.4244	16.4137	16.4270	16.4023
67	16.4777	16.4922	16.4594	16.5013	16.4934	16.4995	16.4925	16.4689	16.4639
68	16.5080	16.5517	16.5519	16.5379	16.5361	16.5370	16.5598	16.5773	16.5546
69	16.5792	16.6317	16.5705	16.6424	16.6275	16.6397	16.5952	16.5884	16.6151
70	16.6469	16.6578	16.6828	16.6425	16.6467	16.6423	16.7041	16.7022	16.6455
71	16.7202	16.7476	16.7077	16.7303	16.7428	16.7246	16.7475	16.7600	16.7169
72	16.7559	16.7979	16.7972	16.8101	16.7860	16.8128	16.8080	16.7851	16.7970
73	16.8342	16.8605	16.8796	16.8726	16.8732	16.8929	16.8870	16.8922	16.8695
74	16.9012	16.9510	16.8872	16.9330	16.9199	16.9096	16.9351	16.9178	16.9073
75	16.9473	17.0419	16.9541	17.0028	16.9804	17.0320	16.9923	17.0201	17.0207
76	17.0631	17.0471	17.0911	17.0828	17.0927	17.0480	17.1125	17.0685	17.0312
77	17.0959	17.1461	17.1616	17.1360	17.1186	17.1522	17.1425	17.1324	17.1588
78	17.2019	17.2404	17.1719	17.2447	17.2483	17.2236	17.2580	17.2525	17.1867
79	17.2304	17.2943	17.2581	17.2920	17.2651	17.3129	17.3301	17.2917	17.2600
80	17.3738	17.4055	17.3854	17.4010	17.4132	17.3763	17.3842	17.4048	17.3971
81	17.4635	17.4502	17.4207	17.4919	17.4898	17.4772	17.4525	17.4360	17.4844
82	17.4670	17.5848	17.5530	17.5357	17.5210	17.5465	17.5982	17.5954	17.5036
83	17.5763	17.6770	17.6296	17.6241	17.6720	17.6282	17.6396	17.6026	17.6160
84	17.7070	17.7192	17.7004	17.7642	17.6982	17.7560	17.7738	17.7903	17.7296
85	17.7959	17.8841	17.7780	17.7947	17.7910	17.8628	17.8283	17.7966	17.8505
86	17.8717	17.9057	17.9423	17.9900	17.9737	17.9141	17.9818	17.9907	17.8846
87	17.9702	18.0132	18.0322	18.0648	18.0595	18.0639	18.0906	18.0324	17.9847
88	18.1258	18.2175	18.1152	18.1511	18.1349	18.1474	18.1539	18.1890	18.1867
89	18.2657	18.3443	18.2515	18.2467	18.3084	18.2991	18.3520	18.2814	18.2122
90	18.3044	18.3701	18.3803	18.4639	18.3729	18.4011	18.3844	18.4274	18.4449
91	18.4619	18.5832	18.4826	18.5222	18.4966	18.5103	18.6044	18.6057	18.5431
92	18.6662	18.6976	18.7117	18.7537	18.7560	18.7616	18.6997	18.6630	18.6680
93	18.8538	18.9571	18.7700	18.8861	18.8020	18.8970	18.9189	18.8232	18.7638
94	18.9272	18.9974	19.0987	19.0592	19.1290	19.0414	19.0636	19.1336	19.1341
95	19.1192	19.3451	19.2714	19.1709	19.1896	19.3655	19.2415	19.3119	19.2532
96	19.5196	19.4649	19.4267	19.6633	19.5997	19.4216	19.6158	19.4873	19.4767
97	19.6563	19.7711	19.6561	19.9071	19.9222	19.8433	19.9388	19.8920	19.6931
98	20.1773	20.2830	20.2986	20.0902	20.0365	20.1641	20.1073	20.1054	20.2779
99	20.3606	20.7707	20.4641	20.7059	20.5946	20.6641	20.9946	20.7349	20.5098
100	23.0731	21.7096	22.5530	21.8410	22.1322	21.9590	21.3619	21.7513	22.3892
101									
102									
103	mean								
104	15.8571	15.8571	15.8571	15.8571	15.8571	15.8571	15.8571	15.8571	15.8571
105	std.dev								
106	1.9070	1.9070	1.9070	1.9070	1.9070	1.9070	1.9070	1.9070	1.9070
107									

Appendix IV: Beta (12, 15, 30)
100 generated values as per 25 runs in column A to Y

	J	K	L	M	N	O	P	Q	R
1	12.3678	12.5817	12.5883	12.2600	12.4393	12.4769	12.6408	12.1314	12.5923
2	12.7641	12.6655	12.6386	12.8031	12.7828	12.7257	12.7703	12.8237	12.6637
3	12.8375	12.8919	12.8461	12.9152	12.8453	12.8821	12.9189	12.8953	12.8408
4	13.0513	13.0240	13.0433	13.0097	13.0952	13.0213	13.1295	13.0389	13.0690
5	13.1248	13.0938	13.0963	13.1282	13.1157	13.1377	13.2161	13.1816	13.1387
6	13.2140	13.2575	13.2404	13.2329	13.2748	13.2041	13.2675	13.1895	13.2207
7	13.2695	13.3303	13.3435	13.2955	13.2934	13.3001	13.3608	13.3237	13.3025
8	13.4221	13.3735	13.3459	13.4168	13.4463	13.3941	13.4636	13.3980	13.4077
9	13.4852	13.4824	13.4681	13.4760	13.5018	13.4644	13.5404	13.5022	13.4920
10	13.5124	13.5261	13.5272	13.5409	13.5400	13.5354	13.5796	13.5236	13.5224
11	13.5759	13.6217	13.6326	13.6230	13.6051	13.5784	13.6540	13.6317	13.6239
12	13.6959	13.6633	13.6401	13.6702	13.7084	13.6954	13.7330	13.6700	13.6666
13	13.7311	13.7295	13.7061	13.7568	13.7218	13.7654	13.8092	13.7175	13.7219
14	13.7977	13.8075	13.8181	13.7896	13.8422	13.7657	13.8248	13.8341	13.8193
15	13.8612	13.8744	13.8755	13.8755	13.8901	13.8630	13.9261	13.8786	13.8868
16	13.9052	13.8998	13.8883	13.9067	13.9110	13.9047	13.9367	13.9113	13.8922
17	13.9659	13.9852	13.9918	13.9752	14.0043	13.9671	14.0068	13.9654	13.9893
18	14.0291	14.0169	14.0008	14.0345	14.0228	14.0287	14.0764	14.0508	14.0172
19	14.0897	14.0784	14.0538	14.0670	14.1204	14.0966	14.1402	14.0722	14.0998
20	14.1226	14.1398	14.1556	14.1589	14.1210	14.1163	14.1533	14.1609	14.1226
21	14.1878	14.2103	14.1920	14.2006	14.2038	14.1892	14.2203	14.1750	14.1700
22	14.2350	14.2176	14.2281	14.2350	14.2455	14.2340	14.2766	14.2674	14.2617
23	14.3105	14.3088	14.3008	14.3068	14.2985	14.2668	14.3089	14.2923	14.3040
24	14.3181	14.3240	14.3250	14.3335	14.3538	14.3609	14.3861	14.3541	14.3325
25	14.3745	14.3928	14.4102	14.3870	14.3812	14.4131	14.4291	14.3754	14.3861
26	14.4536	14.4386	14.4150	14.4518	14.4683	14.4148	14.4592	14.4697	14.4486
27	14.4895	14.4683	14.4734	14.5173	14.5148	14.5068	14.5046	14.4927	14.4853
28	14.5359	14.5591	14.5487	14.5185	14.5300	14.5186	14.5740	14.5484	14.5454
29	14.5996	14.6062	14.5868	14.6081	14.5805	14.5855	14.5977	14.5900	14.5722
30	14.6192	14.6143	14.6290	14.6197	14.6549	14.6326	14.6680	14.6434	14.6511
31	14.6664	14.6811	14.7039	14.6694	14.6795	14.6617	14.6836	14.6797	14.6691
32	14.7443	14.7308	14.7043	14.7493	14.7456	14.7481	14.7676	14.7446	14.7451
33	14.7820	14.7610	14.7795	14.7636	14.7885	14.7811	14.7895	14.7892	14.7592
34	14.8189	14.8401	14.8183	14.8444	14.8242	14.8186	14.8454	14.8240	14.8441
35	14.8951	14.8917	14.8729	14.8821	14.8727	14.8579	14.8826	14.8923	14.8925
36	14.8953	14.8981	14.9144	14.9145	14.9271	14.9310	14.9353	14.9093	14.8993
37	14.9666	14.9569	14.9446	14.9878	14.9670	14.9459	14.9643	14.9495	14.9678
38	15.0123	15.0207	15.0311	14.9965	15.0191	15.0313	15.0359	15.0395	15.0114
39	15.0774	15.0594	15.0427	15.0502	15.0424	15.0704	15.0510	15.0627	15.0545
40	15.0902	15.1061	15.1217	15.1220	15.1301	15.0953	15.1314	15.1138	15.1125
41	15.1454	15.1580	15.1370	15.1649	15.1474	15.1656	15.1706	15.1659	15.1312
42	15.2113	15.1958	15.2164	15.1954	15.2118	15.1890	15.1944	15.1985	15.2238
43	15.2720	15.2388	15.2661	15.2278	15.2577	15.2629	15.2421	15.2606	15.2261
44	15.2744	15.3040	15.2771	15.3215	15.2890	15.2812	15.3063	15.2925	15.3177
45	15.3533	15.3357	15.3316	15.3328	15.3656	15.3441	15.3213	15.3246	15.3630
46	15.3842	15.3975	15.4026	15.4067	15.3698	15.3908	15.4117	15.4184	15.3707
47	15.4335	15.4202	15.4357	15.4346	15.4602	15.4259	15.4396	15.4315	15.4323
48	15.4967	15.5048	15.4912	15.4966	15.4655	15.5015	15.4795	15.5030	15.4931
49	15.5231	15.5512	15.5235	15.5305	15.5407	15.5280	15.5530	15.5316	15.5537
50	15.6017	15.5677	15.5980	15.5945	15.5772	15.5938	15.5542	15.5964	15.5653
51	15.6373	15.6362	15.6241	15.6193	15.6121	15.6221	15.6474	15.6429	15.6178
52	15.6851	15.6793	15.6948	15.7022	15.7009	15.6970	15.6506	15.6814	15.6975
53	15.7437	15.7081	15.7291	15.7394	15.7468	15.7580	15.7008	15.7356	15.7273
54	15.7781	15.8070	15.7895	15.7807	15.7632	15.7603	15.7906	15.7871	15.7867
55	15.8541	15.8151	15.8192	15.8141	15.8079	15.8471	15.8215	15.8205	15.8201

Appendix IV: Beta (12, 15, 30)
100 generated values as per 25 runs in column A to Y

	J	K	L	M	N	O	P	Q	R
56	15.8722	15.9027	15.9035	15.9095	15.9040	15.8754	15.8668	15.9053	15.8970
57	15.9552	15.9114	15.9498	15.9171	15.9221	15.9316	15.9403	15.9598	15.9354
58	15.9773	16.0134	15.9793	16.0132	15.9945	15.9977	15.9470	15.9709	15.9874
59	16.0682	16.0412	16.0645	16.0490	16.0243	16.0553	16.0372	16.0332	16.0574
60	16.0777	16.0946	16.0777	16.0923	16.1020	16.0862	16.0564	16.1096	16.0770
61	16.1359	16.1721	16.1667	16.1567	16.1682	16.1267	16.1361	16.1596	16.1391
62	16.2269	16.1793	16.1920	16.2000	16.1720	16.2319	16.1667	16.1985	16.2107
63	16.2600	16.2578	16.2861	16.2518	16.2270	16.2557	16.2226	16.2565	16.2433
64	16.3254	16.3158	16.2956	16.3271	16.3344	16.3249	16.2958	16.3233	16.3286
65	16.3952	16.4004	16.3580	16.3798	16.3650	16.3492	16.3568	16.3629	16.3593
66	16.4208	16.4028	16.4542	16.4284	16.4232	16.4617	16.3842	16.4460	16.4416
67	16.5201	16.4610	16.5094	16.4942	16.4833	16.4672	16.4342	16.4717	16.4852
68	16.5317	16.5793	16.5385	16.5493	16.5383	16.5813	16.5364	16.5734	16.5510
69	16.6206	16.6083	16.6209	16.6201	16.6182	16.5977	16.5837	16.5906	16.6176
70	16.6780	16.6756	16.6738	16.6685	16.6466	16.6956	16.6235	16.6988	16.6632
71	16.7500	16.7653	16.7357	16.7077	16.7554	16.7456	16.6675	16.7667	16.7069
72	16.8053	16.7741	16.8156	16.8364	16.7629	16.8036	16.7877	16.7770	16.8292
73	16.9005	16.8910	16.8677	16.8860	16.8283	16.8927	16.7966	16.8443	16.8402
74	16.9211	16.9136	16.9501	16.9232	16.9546	16.9225	16.9174	16.9656	16.9622
75	16.9987	16.9736	17.0046	17.0218	16.9708	16.9931	16.9791	17.0237	16.9956
76	17.1049	17.1130	17.0944	17.0660	17.0895	17.1039	17.0032	17.0628	17.0852
77	17.1255	17.1843	17.1293	17.1650	17.1113	17.1723	17.1155	17.1537	17.1154
78	17.2751	17.1942	17.2667	17.2177	17.2412	17.2184	17.1537	17.2275	17.2613
79	17.3361	17.2840	17.2744	17.3267	17.2847	17.3022	17.2567	17.3289	17.3138
80	17.3766	17.4083	17.4356	17.3694	17.3767	17.4030	17.3167	17.3653	17.3733
81	17.5042	17.4294	17.4590	17.4561	17.4269	17.5056	17.3945	17.5016	17.4521
82	17.5446	17.5983	17.5855	17.5746	17.5664	17.5352	17.5035	17.5267	17.5693
83	17.6585	17.6520	17.6975	17.6947	17.6252	17.6916	17.5766	17.6826	17.6228
84	17.7526	17.7356	17.7091	17.6966	17.7259	17.7111	17.6714	17.7057	17.7588
85	17.8645	17.8248	17.8209	17.8814	17.8334	17.8528	17.7922	17.8083	17.7880
86	17.9412	17.9566	17.9823	17.9027	17.9074	17.9440	17.8370	17.9744	17.9899
87	18.0334	18.0209	18.0558	18.0852	17.9809	18.0542	18.0000	18.0552	18.0901
88	18.2133	18.1988	18.1838	18.1316	18.1971	18.1805	18.0495	18.1601	18.1162
89	18.3253	18.2824	18.3321	18.3471	18.2180	18.3403	18.1759	18.2623	18.2391
90	18.4074	18.4240	18.3958	18.3596	18.4450	18.3823	18.3534	18.4440	18.4634
91	18.6361	18.5064	18.5363	18.6129	18.5144	18.5454	18.4742	18.5621	18.5275
92	18.6625	18.7709	18.7654	18.6571	18.7068	18.7492	18.6001	18.7057	18.7382
93	18.8157	18.8157	18.8232	18.8277	18.9376	18.8071	18.8402	18.8754	18.9025
94	19.1805	19.1421	19.1652	19.1298	18.9464	19.1781	18.8858	19.0660	19.0301
95	19.2511	19.2813	19.3113	19.2757	19.1459	19.3957	19.1757	19.1941	19.3815
96	19.5971	19.5222	19.5227	19.5307	19.6197	19.4198	19.3846	19.6222	19.3974
97	19.9418	19.6706	19.7939	19.7719	19.7429	19.9814	19.8256	19.7454	19.7188
98	20.0968	20.4288	20.2774	20.2655	20.2142	20.0405	19.8905	20.2890	20.3284
99	21.0181	20.7388	20.9213	20.7456	20.5281	20.8137	20.3036	20.7471	20.6604
100	21.3215	21.7491	21.4525	21.7358	22.3089	21.6366	23.4558	21.7066	21.9351
101									
102									
103	mean								
104	15.8571	15.8571	15.8571	15.8571	15.8571	15.8571	15.8571	15.8571	15.8571
105	std.dev								
106	1.9070	1.9070	1.9070	1.9070	1.9070	1.9070	1.9070	1.9070	1.9070
107									

Appendix IV: Beta (12, 15, 30)
100 generated values as per 25 runs in column A to Y

	S	T	U	V	W	X	Y
1	12.2793	12.4558	12.5540	12.5937	12.3786	12.6582	12.5790
2	12.7908	12.7375	12.6756	12.7517	12.7795	12.7620	12.6407
3	12.8909	12.9129	12.8723	12.9106	12.8914	12.9184	12.8587
4	13.0189	12.9960	13.0260	13.0855	13.0328	13.1369	13.0285
5	13.1044	13.1423	13.0758	13.1835	13.0867	13.1908	13.1430
6	13.2430	13.2040	13.2604	13.2475	13.2715	13.3002	13.1881
7	13.3298	13.3190	13.3325	13.3679	13.2956	13.3728	13.2982
8	13.3699	13.3793	13.3590	13.4080	13.4165	13.4584	13.3861
9	13.4562	13.4970	13.4873	13.4772	13.4461	13.5517	13.4268
10	13.5490	13.5067	13.5100	13.5979	13.5704	13.5746	13.5638
11	13.6135	13.5761	13.6123	13.6435	13.6388	13.6779	13.5867
12	13.6688	13.7009	13.6624	13.7030	13.6538	13.7161	13.6798
13	13.7122	13.7254	13.7124	13.7620	13.7569	13.8118	13.7507
14	13.8214	13.8076	13.8141	13.8323	13.7887	13.8275	13.7724
15	13.8441	13.8557	13.8407	13.8811	13.8410	13.8787	13.8678
16	13.9286	13.9151	13.9248	13.9459	13.9402	13.9890	13.8925
17	13.9474	13.9515	13.9536	13.9966	13.9860	14.0275	13.9845
18	14.0520	14.0462	14.0394	14.0539	14.0229	14.0608	14.0048
19	14.1050	14.0719	14.0862	14.1143	14.1100	14.1399	14.1017
20	14.1129	14.1438	14.1248	14.1486	14.1147	14.1577	14.1049
21	14.2066	14.2122	14.2001	14.2237	14.2023	14.2326	14.2079
22	14.2215	14.2136	14.2213	14.2451	14.2319	14.2680	14.2094
23	14.3088	14.2772	14.2885	14.2911	14.3162	14.3182	14.3075
24	14.3247	14.3532	14.3382	14.3780	14.3227	14.3803	14.3159
25	14.3786	14.4146	14.3898	14.3983	14.3932	14.4401	14.4053
26	14.4540	14.4155	14.4364	14.4667	14.4440	14.4512	14.4177
27	14.5109	14.4943	14.5067	14.5134	14.4733	14.5264	14.4945
28	14.5193	14.5328	14.5171	14.5447	14.5597	14.5554	14.5262
29	14.5881	14.6083	14.5965	14.6039	14.5945	14.6290	14.5804
30	14.6347	14.6117	14.6200	14.6432	14.6314	14.6391	14.6336
31	14.6712	14.6683	14.6919	14.7009	14.6791	14.6968	14.6571
32	14.7432	14.7433	14.7168	14.7342	14.7378	14.7565	14.7489
33	14.7618	14.7855	14.7886	14.7769	14.8002	14.7726	14.7620
34	14.8425	14.8158	14.8098	14.8438	14.8057	14.8640	14.8346
35	14.8543	14.8518	14.8772	14.8737	14.8793	14.8827	14.8778
36	14.9390	14.9384	14.9104	14.9322	14.9151	14.9364	14.9085
37	14.9646	14.9495	14.9530	14.9860	14.9768	14.9909	14.9620
38	15.0171	15.0290	15.0230	15.0043	15.0051	15.0101	15.0130
39	15.0499	15.0380	15.0441	15.0567	15.0646	15.0809	15.0458
40	15.1202	15.1288	15.1205	15.1182	15.1051	15.1021	15.1181
41	15.1792	15.1430	15.1668	15.1527	15.1464	15.1677	15.1387
42	15.1797	15.2125	15.1867	15.2070	15.2113	15.1976	15.2144
43	15.2349	15.2442	15.2588	15.2533	15.2713	15.2683	15.2414
44	15.3135	15.3006	15.2843	15.2921	15.2752	15.2799	15.3018
45	15.3521	15.3288	15.3612	15.3442	15.3482	15.3301	15.3611
46	15.3870	15.4067	15.3728	15.3881	15.3884	15.4025	15.3733
47	15.4508	15.4233	15.4607	15.4246	15.4385	15.4261	15.4522
48	15.4808	15.5044	15.4659	15.4961	15.4898	15.4923	15.4751
49	15.5580	15.5360	15.5567	15.5372	15.5247	15.5504	15.5397
50	15.5678	15.5860	15.5644	15.5737	15.5972	15.5556	15.5824
51	15.6155	15.6182	15.6294	15.6219	15.6328	15.6366	15.6440
52	15.7075	15.7009	15.6889	15.6821	15.6856	15.6599	15.6758
53	15.7277	15.7526	15.7549	15.7088	15.7246	15.7306	15.7586
54	15.7949	15.7655	15.7628	15.7910	15.7925	15.7583	15.7609
55	15.8456	15.8327	15.8564	15.8253	15.8440	15.8234	15.8397

Appendix IV: Beta (12, 15, 30)
100 generated values as per 25 runs in column A to Y

	S	T	U	V	W	X	Y
56	15.8807	15.8894	15.8655	15.8738	15.8761	15.8626	15.8844
57	15.9199	15.9564	15.9125	15.9279	15.9373	15.9348	15.9170
58	16.0138	15.9714	16.0171	15.9733	15.9886	15.9500	16.0151
59	16.0404	16.0514	16.0428	16.0319	16.0269	16.0338	16.0561
60	16.1048	16.0894	16.0983	16.0774	16.1106	16.0568	16.0881
61	16.1708	16.1748	16.1777	16.1525	16.1642	16.1333	16.1571
62	16.1907	16.1821	16.1798	16.1686	16.1888	16.1661	16.2039
63	16.2405	16.2363	16.2403	16.2694	16.2585	16.2450	16.2379
64	16.3445	16.3441	16.3408	16.2699	16.3163	16.2694	16.3473
65	16.3733	16.4037	16.3786	16.3672	16.3946	16.3251	16.3550
66	16.4410	16.4058	16.4321	16.3975	16.4095	16.4117	16.4600
67	16.4660	16.5118	16.5073	16.4651	16.4687	16.4366	16.4989
68	16.5861	16.5329	16.5390	16.5310	16.5718	16.5291	16.5527
69	16.6207	16.6425	16.6333	16.5708	16.5908	16.5952	16.6081
70	16.6755	16.6486	16.6597	16.6662	16.6940	16.6070	16.6901
71	16.7389	16.7430	16.7201	16.6895	16.7063	16.7190	16.7127
72	16.8136	16.8042	16.8292	16.7981	16.8332	16.7307	16.8424
73	16.9049	16.8479	16.8524	16.8608	16.8590	16.8420	16.8878
74	16.9135	16.9664	16.9642	16.8872	16.9458	16.8645	16.9340
75	17.0455	17.0248	17.0170	17.0014	17.0051	16.9644	17.0218
76	17.0512	17.0674	17.0788	17.0194	17.0788	17.0122	17.0807
77	17.1551	17.1601	17.1407	17.1029	17.1798	17.0653	17.1484
78	17.2390	17.2279	17.2519	17.2099	17.1978	17.1991	17.2512
79	17.3386	17.2995	17.2826	17.2477	17.3403	17.2090	17.2845
80	17.3696	17.4027	17.4241	17.3720	17.3505	17.3580	17.4303
81	17.5192	17.4417	17.4709	17.4032	17.4898	17.4120	17.4872
82	17.5245	17.5961	17.5707	17.5449	17.5354	17.4779	17.5630
83	17.6673	17.6752	17.6856	17.6033	17.6335	17.5342	17.6786
84	17.7381	17.7238	17.7180	17.6990	17.7520	17.7063	17.7343
85	17.8417	17.8816	17.8859	17.8127	17.8471	17.8086	17.8889
86	17.9577	17.9112	17.9120	17.8755	17.9309	17.8111	17.9190
87	18.0216	18.0474	18.0493	18.0060	17.9970	17.9423	18.0237
88	18.2188	18.1833	18.1872	18.1100	18.2218	18.1020	18.2266
89	18.2683	18.2628	18.2987	18.2652	18.2351	18.1710	18.3466
90	18.4617	18.4602	18.4274	18.3293	18.4723	18.3471	18.3892
91	18.5067	18.5935	18.4981	18.4596	18.5895	18.5235	18.5074
92	18.7942	18.6901	18.8038	18.6969	18.6741	18.5357	18.8068
93	18.8661	18.8397	18.8362	18.7578	18.8863	18.7717	18.9156
94	19.1068	19.1306	19.1441	19.0669	19.0533	18.9442	19.0667
95	19.3514	19.3939	19.2032	19.1597	19.2335	19.0849	19.4013
96	19.4711	19.4145	19.6499	19.5039	19.5698	19.4763	19.4327
97	19.8158	19.9380	19.9135	19.6414	19.9576	19.6185	19.8550
98	20.2398	20.0770	20.1168	20.2523	20.0285	20.1261	20.2119
99	20.9177	20.7714	20.8520	20.3988	20.6751	20.2939	21.0147
100	21.4433	21.7163	21.5607	22.8589	21.9082	23.4830	21.3355
101							
102							
103	mean						
104	15.8571	15.8571	15.8571	15.8571	15.8571	15.8571	15.8571
105	std.dev						
106	1.9070	1.9070	1.9070	1.9070	1.9070	1.9070	1.9070
107							

Appendix V

Miscellaneous tests of mean values, standard deviations and Pearson correlation coefficients for different combinations of distributions

Note that the expected values and Variances (and hence the standard deviations) of the general Uniform, Triangular, Normal and Beta distributions are stated in sections 6.2.2 to 6.2.5 respectively.

Note also that the reported precision of the calculated correlation coefficient when the target value is 0.7 is sometimes better for samples of size 100 than for samples of size 500. This apparent anomaly arises because only a maximum of four iterations in the swapping routine was allowed. The anomaly disappears when, as previously discussed, this maximum iteration limit is increased to, say, 20.

Expected correlation	-0.6

Mean value of computed correlation of 25 runs		Sample size		
Distribution combination		10	100	500
B(12,15,30)	B(12,15,30)	-0.5994	-0.6000	-0.6000
B(12,15,30)	N(100,36)	-0.6002	-0.6000	-0.6000
B(12,15,30)	T(10,15,40)	-0.6005	-0.6000	-0.6000
B(12,15,30)	U(18,38)	-0.5996	-0.6000	-0.6000
B(12,15,30)	B(12,25,30)	-0.6005	-0.6000	-0.6000
B(12,15,30)	T(10,30,40)	-0.6011	-0.6000	-0.6000
N(100,36)	B(12,15,30)	-0.5985	-0.6000	-0.6000
N(100,36)	N(100,36)	-0.5991	-0.6000	-0.6000
N(100,36)	T(10,15,40)	-0.5996	-0.6000	-0.6000
N(100,36)	U(18,38)	-0.5998	-0.6000	-0.6000
N(100,36)	B(12,25,30)	-0.5994	-0.6000	-0.6000
N(100,36)	T(10,30,40)	-0.6002	-0.6000	-0.6000
T(10,15,40)	B(12,15,30)	-0.5992	-0.6000	-0.6000
T(10,15,40)	N(100,36)	-0.6009	-0.6000	-0.6000
T(10,15,40)	T(10,15,40)	-0.6000	-0.6000	-0.5999
T(10,15,40)	U(18,38)	-0.6009	-0.6000	-0.6000
T(10,15,40)	B(12,25,30)	-0.6003	-0.6000	-0.6000
T(10,15,40)	T(10,30,40)	-0.5996	-0.6000	-0.6000
U(18,38)	B(12,15,30)	-0.5997	-0.6000	-0.6000
U(18,38)	N(100,36)	-0.599	-0.6000	-0.6000
U(18,38)	T(10,15,40)	-0.5995	-0.6000	-0.6000
U(18,38)	U(18,38)	-0.5994	-0.6000	-0.6000
U(18,38)	B(12,25,30)	-0.5986	-0.6000	-0.6000
U(18,38)	T(10,30,40)	-0.5997	-0.6000	-0.6000
B(12,25,30)	B(12,15,30)	-0.5995	-0.6000	-0.6000
B(12,25,30)	N(100,36)	-0.5994	-0.6000	-0.6000
B(12,25,30)	T(10,15,40)	-0.5986	-0.6000	-0.6000
B(12,25,30)	U(18,38)	-0.6000	-0.6000	-0.6000
B(12,25,30)	B(12,25,30)	-0.6001	-0.6000	-0.6000
B(12,25,30)	T(10,30,40)	-0.5995	-0.6000	-0.6000
T(10,30,40)	B(12,15,30)	-0.5987	-0.6000	-0.6000
T(10,30,40)	N(100,36)	-0.5995	-0.6000	-0.6000
T(10,30,40)	U(18,38)	-0.5993	-0.6000	-0.6000
T(10,30,40)	B(12,25,30)	-0.6004	-0.6000	-0.6000
T(10,30,40)	T(10,15,40)	-0.5988	-0.6000	-0.6000
T(10,30,40)	T(10,30,40)	-0.5997	-0.6000	-0.6000

Expected correlation	-0.4

Mean value of computed correlation of 25 runs		Sample size		
Distribution combination		10	100	500
B(12,15,30)	B(12,15,30)	-0.3999	-0.4000	-0.4000
B(12,15,30)	N(100,36)	-0.3994	-0.4000	-0.4000
B(12,15,30)	T(10,15,40)	-0.3990	-0.4000	-0.4000
B(12,15,30)	U(18,38)	-0.3999	-0.4000	-0.4000
B(12,15,30)	B(12,25,30)	-0.3997	-0.4000	-0.4000
B(12,15,30)	T(10,30,40)	-0.3996	-0.4000	-0.4000
N(100,36)	B(12,15,30)	-0.3998	-0.4000	-0.4000
N(100,36)	N(100,36)	-0.3994	-0.4000	-0.4000
N(100,36)	T(10,15,40)	-0.4001	-0.4000	-0.4000
N(100,36)	U(18,38)	-0.3994	-0.4000	-0.4000
N(100,36)	B(12,25,30)	-0.3982	-0.4000	-0.4000
N(100,36)	T(10,30,40)	-0.3994	-0.4000	-0.4000
T(10,15,40)	B(12,15,30)	-0.4001	-0.4000	-0.4000
T(10,15,40)	N(100,36)	-0.3995	-0.4000	-0.4000
T(10,15,40)	T(10,15,40)	-0.3993	-0.4000	-0.4000
T(10,15,40)	U(18,38)	-0.4006	-0.4000	-0.4000
T(10,15,40)	B(12,25,30)	-0.3994	-0.4000	-0.4000
T(10,15,40)	T(10,30,40)	-0.3992	-0.4000	-0.4000
U(18,38)	B(12,15,30)	-0.3997	-0.4000	-0.4000
U(18,38)	N(100,36)	-0.3993	-0.4000	-0.4000
U(18,38)	T(10,15,40)	-0.4008	-0.4000	-0.4000
U(18,38)	U(18,38)	-0.3998	-0.4000	-0.4000
U(18,38)	B(12,25,30)	-0.4001	-0.4000	-0.4000
U(18,38)	T(10,30,40)	-0.4003	-0.4000	-0.4000
B(12,25,30)	B(12,15,30)	-0.4007	-0.4000	-0.4000
B(12,25,30)	N(100,36)	-0.4002	-0.4000	-0.4000
B(12,25,30)	T(10,15,40)	-0.4001	-0.4000	-0.4000
B(12,25,30)	U(18,38)	-0.4006	-0.4000	-0.4000
B(12,25,30)	B(12,25,30)	-0.3998	-0.4000	-0.4000
B(12,25,30)	T(10,30,40)	-0.3989	-0.4000	-0.4000
T(10,30,40)	B(12,15,30)	-0.3998	-0.4000	-0.4000
T(10,30,40)	N(100,36)	-0.4007	-0.4000	-0.4000
T(10,30,40)	U(18,38)	-0.3998	-0.4000	-0.4000
T(10,30,40)	B(12,25,30)	-0.4007	-0.4000	-0.4000
T(10,30,40)	T(10,15,40)	-0.4001	-0.4000	-0.4000
T(10,30,40)	T(10,30,40)	-0.3989	-0.4000	-0.4000

Expected correlation	-0.2

Mean value of computed correlation of 25 runs		Sample size		
Distribution combination		10	100	500
B(12,15,30)	B(12,15,30)	-0.1998	-0.2000	-0.2000
B(12,15,30)	N(100,36)	-0.2004	-0.2000	-0.2000
B(12,15,30)	T(10,15,40)	-0.2012	-0.2000	-0.2000
B(12,15,30)	U(18,38)	-0.2006	-0.2000	-0.2000
B(12,15,30)	B(12,25,30)	-0.1999	-0.2000	-0.2000
B(12,15,30)	T(10,30,40)	-0.1999	-0.2000	-0.2000
N(100,36)	B(12,15,30)	-0.2007	-0.2000	-0.2000
N(100,36)	N(100,36)	-0.1991	-0.2000	-0.2000
N(100,36)	T(10,15,40)	-0.1990	-0.2000	-0.2000
N(100,36)	U(18,38)	-0.2001	-0.2000	-0.2000
N(100,36)	B(12,25,30)	-0.1998	-0.2000	-0.2000
N(100,36)	T(10,30,40)	-0.2007	-0.2000	-0.2000
T(10,15,40)	B(12,15,30)	-0.2013	-0.2000	-0.2000
T(10,15,40)	N(100,36)	-0.2008	-0.2000	-0.2000
T(10,15,40)	T(10,15,40)	-0.2005	-0.2000	-0.2000
T(10,15,40)	U(18,38)	-0.2006	-0.2000	-0.2000
T(10,15,40)	B(12,25,30)	-0.2008	-0.2000	-0.2000
T(10,15,40)	T(10,30,40)	-0.2003	-0.2000	-0.2000
U(18,38)	B(12,15,30)	-0.2002	-0.2000	-0.2000
U(18,38)	N(100,36)	-0.2004	-0.2000	-0.2000
U(18,38)	T(10,15,40)	-0.2001	-0.2000	-0.2000
U(18,38)	U(18,38)	-0.2001	-0.2000	-0.2000
U(18,38)	B(12,25,30)	-0.1997	-0.2000	-0.2000
U(18,38)	T(10,30,40)	-0.2008	-0.2000	-0.2000
B(12,25,30)	B(12,15,30)	-0.2006	-0.2000	-0.2000
B(12,25,30)	N(100,36)	-0.2003	-0.2000	-0.2000
B(12,25,30)	T(10,15,40)	-0.2006	-0.2000	-0.2000
B(12,25,30)	U(18,38)	-0.2014	-0.2000	-0.2000
B(12,25,30)	B(12,25,30)	-0.2012	-0.2000	-0.2000
B(12,25,30)	T(10,30,40)	-0.2002	-0.2000	-0.2000
T(10,30,40)	B(12,15,30)	-0.2000	-0.2000	-0.2000
T(10,30,40)	N(100,36)	-0.2007	-0.2000	-0.2000
T(10,30,40)	U(18,38)	-0.2001	-0.2000	-0.2000
T(10,30,40)	B(12,25,30)	-0.1998	-0.2000	-0.2000
T(10,30,40)	T(10,15,40)	-0.1998	-0.2000	-0.2000
T(10,30,40)	T(10,30,40)	-0.1997	-0.2000	-0.2000

Expected correlation	0.3

Mean value of computed correlation of 25 runs		Sample size		
Distribution combination		10	100	500
B(12,15,30)	B(12,15,30)	0.2998	0.3000	0.3000
B(12,15,30)	N(100,36)	0.2996	0.3000	0.3000
B(12,15,30)	T(10,15,40)	0.3001	0.3000	0.3000
B(12,15,30)	U(18,38)	0.3007	0.3000	0.3000
B(12,15,30)	B(12,25,30)	0.3004	0.3000	0.3000
B(12,15,30)	T(10,30,40)	0.3019	0.3000	0.3000
N(100,36)	B(12,15,30)	0.3002	0.3000	0.3000
N(100,36)	N(100,36)	0.3003	0.3000	0.3000
N(100,36)	T(10,15,40)	0.2990	0.3000	0.3000
N(100,36)	U(18,38)	0.3008	0.3000	0.3000
N(100,36)	B(12,25,30)	0.2998	0.3000	0.3000
N(100,36)	T(10,30,40)	0.2995	0.3000	0.3000
T(10,15,40)	B(12,15,30)	0.3005	0.3000	0.3000
T(10,15,40)	N(100,36)	0.2994	0.3000	0.3000
T(10,15,40)	T(10,15,40)	0.3005	0.3000	0.3000
T(10,15,40)	U(18,38)	0.3009	0.3000	0.3000
T(10,15,40)	B(12,25,30)	0.3004	0.3000	0.3000
T(10,15,40)	T(10,30,40)	0.2993	0.3000	0.3000
U(18,38)	B(12,15,30)	0.3004	0.3000	0.3000
U(18,38)	N(100,36)	0.2999	0.3000	0.3000
U(18,38)	T(10,15,40)	0.2992	0.3000	0.3000
U(18,38)	U(18,38)	0.3001	0.3000	0.3000
U(18,38)	B(12,25,30)	0.2997	0.3000	0.3000
U(18,38)	T(10,30,40)	0.3002	0.3000	0.3000
B(12,25,30)	B(12,15,30)	0.300	0.3000	0.3000
B(12,25,30)	N(100,36)	0.3009	0.3000	0.3000
B(12,25,30)	T(10,15,40)	0.3003	0.3000	0.3000
B(12,25,30)	U(18,38)	0.3007	0.3000	0.3000
B(12,25,30)	B(12,25,30)	0.2994	0.3000	0.3000
B(12,25,30)	T(10,30,40)	0.3004	0.3000	0.3000
T(10,30,40)	B(12,15,30)	0.2993	0.3000	0.3000
T(10,30,40)	N(100,36)	0.2996	0.3000	0.3000
T(10,30,40)	U(18,38)	0.3002	0.3000	0.3000
T(10,30,40)	B(12,25,30)	0.2991	0.3000	0.3000
T(10,30,40)	T(10,15,40)	0.2996	0.3000	0.3000
T(10,30,40)	T(10,30,40)	0.3001	0.3000	0.3000

Expected correlation	0.5

Mean value of computed correlation of 25 runs		Sample size		
Distribution combination		10	100	500
B(12,15,30)	B(12,15,30)	0.5001	0.5000	0.5000
B(12,15,30)	N(100,36)	0.4993	0.5000	0.5000
B(12,15,30)	T(10,15,40)	0.5000	0.5000	0.5000
B(12,15,30)	U(18,38)	0.4999	0.5000	0.5000
B(12,15,30)	B(12,25,30)	0.5000	0.5000	0.5000
B(12,15,30)	T(10,30,40)	0.5006	0.5000	0.5000
N(100,36)	B(12,15,30)	0.4998	0.5000	0.5000
N(100,36)	N(100,36)	0.5003	0.5000	0.5000
N(100,36)	T(10,15,40)	0.4997	0.5000	0.5000
N(100,36)	U(18,38)	0.5003	0.5000	0.5000
N(100,36)	B(12,25,30)	0.501	0.5000	0.5000
N(100,36)	T(10,30,40)	0.5005	0.5000	0.5000
T(10,15,40)	B(12,15,30)	0.4991	0.5000	0.5000
T(10,15,40)	N(100,36)	0.4998	0.5000	0.5000
T(10,15,40)	T(10,15,40)	0.4995	0.5000	0.5000
T(10,15,40)	U(18,38)	0.4999	0.5000	0.5000
T(10,15,40)	B(12,25,30)	0.4988	0.5000	0.5000
T(10,15,40)	T(10,30,40)	0.5000	0.5000	0.5000
U(18,38)	B(12,15,30)	0.5001	0.5000	0.5000
U(18,38)	N(100,36)	0.4997	0.5000	0.5000
U(18,38)	T(10,15,40)	0.5000	0.5000	0.5000
U(18,38)	U(18,38)	0.4992	0.5000	0.5000
U(18,38)	B(12,25,30)	0.5002	0.5000	0.5000
U(18,38)	T(10,30,40)	0.4993	0.5000	0.5000
B(12,25,30)	B(12,15,30)	0.5002	0.5000	0.5000
B(12,25,30)	N(100,36)	0.5005	0.5000	0.5000
B(12,25,30)	T(10,15,40)	0.4999	0.5000	0.5000
B(12,25,30)	U(18,38)	0.5002	0.5000	0.5000
B(12,25,30)	B(12,25,30)	0.5009	0.5000	0.5000
B(12,25,30)	T(10,30,40)	0.5001	0.5000	0.5000
T(10,30,40)	B(12,15,30)	0.5016	0.5000	0.5000
T(10,30,40)	N(100,36)	0.5000	0.5000	0.5000
T(10,30,40)	U(18,38)	0.4999	0.5000	0.5000
T(10,30,40)	B(12,25,30)	0.4995	0.5000	0.5000
T(10,30,40)	T(10,15,40)	0.5002	0.5000	0.5000
T(10,30,40)	T(10,30,40)	0.5007	0.5000	0.5000

Expected correlation	0.7

Mean value of computed correlation of 25 runs		Sample size		
Distribution combination		10	100	500
B(12,15,30)	B(12,15,30)	0.7002	0.7000	0.7000
B(12,15,30)	N(100,36)	0.6998	0.7000	0.7000
B(12,15,30)	T(10,15,40)	0.7004	0.7000	0.7000
B(12,15,30)	U(18,38)	0.7001	0.7000	0.6993
B(12,15,30)	B(12,25,30)	0.701	0.7000	0.6999
B(12,15,30)	T(10,30,40)	0.7001	0.7000	0.6982
N(100,36)	B(12,15,30)	0.7002	0.7000	0.6996
N(100,36)	N(100,36)	0.6992	0.7000	0.6997
N(100,36)	T(10,15,40)	0.6998	0.7000	0.6963
N(100,36)	U(18,38)	0.7000	0.7000	0.6955
N(100,36)	B(12,25,30)	0.6998	0.7000	0.6998
N(100,36)	T(10,30,40)	0.7000	0.7000	0.6994
T(10,15,40)	B(12,15,30)	0.6989	0.7000	0.7000
T(10,15,40)	N(100,36)	0.7002	0.7000	0.7000
T(10,15,40)	T(10,15,40)	0.6997	0.7000	0.7000
T(10,15,40)	U(18,38)	0.6994	0.7000	0.7000
T(10,15,40)	B(12,25,30)	0.7001	0.7000	0.7000
T(10,15,40)	T(10,30,40)	0.7005	0.7000	0.6968
U(18,38)	B(12,15,30)	0.6999	0.7000	0.7000
U(18,38)	N(100,36)	0.7003	0.7000	0.7000
U(18,38)	T(10,15,40)	0.6996	0.7000	0.7000
U(18,38)	U(18,38)	0.6998	0.7000	0.7000
U(18,38)	B(12,25,30)	0.7002	0.7000	0.7000
U(18,38)	T(10,30,40)	0.7006	0.7000	0.7000
B(12,25,30)	B(12,15,30)	0.7000	0.7000	0.7000
B(12,25,30)	N(100,36)	0.7001	0.7000	0.7000
B(12,25,30)	T(10,15,40)	0.7001	0.7000	0.6993
B(12,25,30)	U(18,38)	0.6995	0.7000	0.7000
B(12,25,30)	B(12,25,30)	0.6986	0.7000	0.7000
B(12,25,30)	T(10,30,40)	0.6997	0.7000	0.7000
T(10,30,40)	B(12,15,30)	0.6993	0.7000	0.6985
T(10,30,40)	N(100,36)	0.6993	0.7000	0.7000
T(10,30,40)	U(18,38)	0.6993	0.7000	0.7000
T(10,30,40)	B(12,25,30)	0.7004	0.7000	0.7000
T(10,30,40)	T(10,15,40)	0.6993	0.7000	0.6932
T(10,30,40)	T(10,30,40)	0.6999	0.7000	0.7000

Correlation of Normal(100,36) with Uniform[18,38) Computed sample means of 25 runs			
Correlation coefficient	Sample size	Mean value of computed mean values	
-0.6	10	100.0000	28.0000
-0.6	100	100.0000	28.0000
-0.6	500	100.0000	28.0000
-0.4	10	100.0000	28.0000
-0.4	100	100.0000	28.0000
-0.4	500	100.0000	28.0000
-0.2	10	100.0000	28.0000
-0.2	100	100.0000	28.0000
-0.2	500	100.0000	28.0000
0.3	10	100.0000	28.0000
0.3	100	100.0000	28.0000
0.3	500	100.0000	28.0000
0.5	10	100.0000	28.0000
0.5	100	100.0000	28.0000
0.5	500	100.0000	28.0000
0.7	10	100.0000	28.0000
0.7	100	100.0000	28.0000
0.7	500	100.0000	28.0000
Expected mean values of N(100,36) and U[18,38) are 100 and 28 exactly			

Correlation of Triangular(10,15,40) with Triangular(10,30,40) Computed sample means of 25 runs			
Correlation coefficient	Sample size	Mean value of computed mean values	
-0.6	10	21.6667	26.6667
-0.6	100	21.6667	26.6667
-0.6	500	21.6667	26.6667
-0.4	10	21.6667	26.6667
-0.4	100	21.6667	26.6667
-0.4	500	21.6667	26.6667
-0.2	10	21.6667	26.6667
-0.2	100	21.6667	26.6667
-0.2	500	21.6667	26.6667
0.3	10	21.6667	26.6667
0.3	100	21.6667	26.6667
0.3	500	21.6667	26.6667
0.5	10	21.6667	26.6667
0.5	100	21.6667	26.6667
0.5	500	21.6667	26.6667
0.7	10	21.6667	26.6667
0.7	100	21.6667	26.6667
0.7	500	21.6667	26.6667
Expected mean values of T(10,15,40) and T(10,30,40) are 21.6667 and 26.6667 (4 d.p.s)			

Correlation of Beta(12,15,30) with Beta(12,25,30) Computed sample means of 25 runs			
Correlation coefficient	Sample size	Mean value of computed mean values	
-0.6	10	15.8571	24.1304
-0.6	100	15.8571	24.1304
-0.6	500	15.8571	24.1304
-0.4	10	15.8571	24.1304
-0.4	100	15.8571	24.1304
-0.4	500	15.8571	24.1304
-0.2	10	15.8571	24.1304
-0.2	100	15.8571	24.1304
-0.2	500	15.8571	24.1304
0.3	10	15.8571	24.1304
0.3	100	15.8571	24.1304
0.3	500	15.8571	24.1304
0.5	10	15.8571	24.1304
0.5	100	15.8571	24.1304
0.5	500	15.8571	24.1304
0.7	10	15.8571	24.1304
0.7	100	15.8571	24.1304
0.7	500	15.8571	24.1304
Expected mean values of B(12,15,30) and B(12,25,30) are 15.8571 and 24.1304 (4 d.p.s)			

Correlation of Normal(100,36) with Uniform(18,38) Mean value of computed standard deviation values of 25 runs			
Correlation coefficient	Sample size	Mean value of computed standard deviation values	
-0.6	10	6.0000	5.7735
-0.6	100	6.0000	5.7735
-0.6	500	6.0000	5.7735
-0.4	10	6.0000	5.7735
-0.4	100	6.0000	5.7735
-0.4	500	6.0000	5.7735
-0.2	10	6.0000	5.7735
-0.2	100	6.0000	5.7735
-0.2	500	6.0000	5.7735
0.3	10	6.0000	5.7735
0.3	100	6.0000	5.7735
0.3	500	6.0000	5.7735
0.5	10	6.0000	5.7735
0.5	100	6.0000	5.7735
0.5	500	6.0000	5.7735
0.7	10	6.0000	5.7735
0.7	100	6.0000	5.7735
0.7	500	6.0000	5.7735
Expected s.d. values of N(100,36) and U(18,38) are 6 exactly and 5.7735 (4 d.p.s)			

Correlation of Triangular(10,15,40) with Triangular(10,30,40) Mean value of computed standard deviation values of 25 runs			
Correlation coefficient	Sample size	Mean value of computed standard deviation values	
-0.6	10	6.5617	6.2361
-0.6	100	6.5617	6.2361
-0.6	500	6.5617	6.2361
-0.4	10	6.5617	6.2361
-0.4	100	6.5617	6.2361
-0.4	500	6.5617	6.2361
-0.2	10	6.5617	6.2361
-0.2	100	6.5617	6.2361
-0.2	500	6.5617	6.2361
0.3	10	6.5617	6.2361
0.3	100	6.5617	6.2361
0.3	500	6.5617	6.2361
0.5	10	6.5617	6.2361
0.5	100	6.5617	6.2361
0.5	500	6.5617	6.2361
0.7	10	6.5617	6.2361
0.7	100	6.5617	6.2361
0.7	500	6.5617	6.2361
Expected s.d. values of T(10,15,40) and T(10,30,40) are 6.5617 and 6.2361 (4 d.p.s)			

Correlation of Beta(12,15,30) with Beta(12,25,30) Mean value of computed standard deviation values of 25 runs			
Correlation coefficient	Sample size	Mean value of computed standard deviation values	
-0.6	10	1.9070	2.6421
-0.6	100	1.9070	2.6421
-0.6	500	1.9070	2.6421
-0.4	10	1.9070	2.6421
-0.4	100	1.9070	2.6421
-0.4	500	1.9070	2.6421
-0.2	10	1.9070	2.6421
-0.2	100	1.9070	2.6421
-0.2	500	1.9070	2.6421
0.3	10	1.9070	2.6421
0.3	100	1.9070	2.6421
0.3	500	1.9070	2.6421
0.5	10	1.9070	2.6421
0.5	100	1.9070	2.6421
0.5	500	1.9070	2.6421
0.7	10	1.9070	2.6421
0.7	100	1.9070	2.6421
0.7	500	1.9070	2.6421
Expected s.d. values of B(12,15,30) and B(12,25,30) are 1.9070 and 2.6421 (4 d.p.s)			

Appendix VI

χ^2 tests of 25 runs of size 100 from each distribution

Within this appendix samples from each of the four general distributions used in the RCM are tested for goodness-of-fit using the chi-square distributions.

For example in the third case, in the first run of 25 of generating samples of size 100 from N(100,36), the distribution of the 100 observed values are compared with the frequencies within which they would be expected to fall into 10 equi-probable mutually exhaustive classes, so that the expected frequencies are always all equal to 10 exactly. Here there are 10 classes and 7 degrees of freedom, so that the critical value at the 99.5 percent significance level is 0.989 (from Murdoch, J & Barnes, J.A. (1970). *Statistical Tables for Science, Engineering, Management and Business Studies*. Macmillan, 2nd Ed.). The largest value recorded here is for run 11, with $\chi^2 = 0.6061$, so that even this worst-case value is more critical than the tabulated value, the probability that this sample is not from N(100,36) being less than 1 in 200.

Similar conclusions can be drawn from the results for the other three general distributions.

χ^2 test of 25 runs of size 100 from Uniform (18, 38): observed frequencies

Run	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	No. of 10	No. of 9	No. of 11	χ^2
1	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
2	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
3	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
4	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
5	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
6	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
7	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
8	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
9	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
10	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
11	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
12	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
13	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
14	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
15	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
16	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
17	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
18	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
19	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
20	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
21	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
22	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
23	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
24	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
25	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000

χ^2 test of 25 runs of size 100 from Triangular (10, 30 , 40): observed frequencies

Run	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	No. of 10	No. of 9	No. of 11	χ^2
1	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
2	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
3	10	10	10	11	9	10	10	10	10	10	8	1	1	0.2020
4	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
5	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
6	10	10	10	10	10	10	10	9	11	10	8	1	1	0.2020
7	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
8	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
9	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
10	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
11	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
12	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
13	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
14	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
15	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
16	10	10	10	10	10	10	10	10	9	11	8	1	1	0.2020
17	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
18	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
19	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
20	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
21	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
22	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
23	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
24	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
25	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000

χ^2 test of 25 runs of size 100 from Normal (100, 36): observed frequencies

Run	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	No. of 10	No. of 9	No. of 11	χ^2
1	10	10	10	10	10	11	10	10	10	9	8	1	1	0.2020
2	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
3	10	11	9	11	9	10	10	9	10	11	4	3	3	0.6061
4	10	10	10	10	10	10	10	9	11	10	8	1	1	0.2020
5	10	9	11	10	10	10	10	10	10	10	8	1	1	0.2020
6	10	10	10	10	10	10	10	9	11	10	8	1	1	0.2020
7	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
8	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
9	11	9	10	10	10	10	10	10	10	10	8	1	1	0.2020
10	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
11	9	11	9	11	10	11	10	9	10	10	4	3	3	0.6061
12	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
13	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
14	10	10	10	10	10	10	10	10	9	11	8	1	1	0.2020
15	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
16	10	10	10	11	9	10	10	10	9	11	6	2	2	0.4040
17	11	9	10	10	10	10	10	10	10	10	8	1	1	0.2020
18	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
19	10	11	9	10	10	10	10	10	10	10	8	1	1	0.2020
20	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
21	11	9	10	10	10	10	10	10	10	10	8	1	1	0.2020
22	10	10	11	9	10	10	10	9	11	10	6	2	2	0.4040
23	10	11	9	10	10	10	10	10	10	10	8	1	1	0.2020
24	10	9	11	10	10	10	10	10	10	10	8	1	1	0.2020
25	10	10	11	9	10	10	9	11	10	10	6	2	2	0.4040

χ^2 test of 25 runs of size 100 from Beta (12, 15, 30): observed frequencies

Run	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	No. of 10	No. of 9	No. of 11	χ^2
1	9	11	10	10	10	10	10	10	11	9	6	2	2	0.4040
2	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
3	10	9	10	11	10	10	10	10	10	10	8	1	1	0.2020
4	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
5	10	11	9	10	10	10	10	10	10	10	8	1	1	0.2020
6	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
7	10	11	9	10	10	10	9	11	10	10	6	2	2	0.4040
8	11	9	10	10	10	10	9	11	10	10	6	2	2	0.4040
9	10	10	10	10	11	9	10	10	10	10	8	1	1	0.2020
10	11	9	10	10	10	10	10	10	10	10	8	1	1	0.2020
11	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
12	10	10	10	10	10	10	10	9	11	10	8	1	1	0.2020
13	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
14	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
15	11	9	11	9	10	10	10	10	10	10	6	2	2	0.4040
16	10	10	9	11	10	10	11	10	10	9	6	2	2	0.4040
17	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
18	10	10	10	11	9	10	10	10	10	10	8	1	1	0.2020
19	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
20	11	9	10	10	10	10	10	10	10	10	8	1	1	0.2020
21	10	10	10	10	10	10	10	9	11	10	8	1	1	0.2020
22	9	11	10	10	10	10	11	10	10	9	6	2	2	0.4040
23	10	10	10	10	10	10	10	10	10	10	10	0	0	0.0000
24	10	10	10	10	10	10	10	11	9	10	8	1	1	0.2020
25	10	10	11	9	10	10	10	9	11	10	6	2	2	0.4040

Appendix VII

Miscellaneous tests of skewness and kurtosis for different distributions

Within this appendix, the following results are utilised:

(1) The expected values of the skewness of the general Uniform, Triangular, Normal and Beta distributions are 0, $-1\{(a+m-2b)(a-2m+b)(-2a+m+b)\}/270\sigma^3$, 0, and $2[(w-v)/(v+w+2)]\,[(v+w+1)/vw]^{\frac{1}{2}}$ respectively.

(2) Their corresponding expected values of the kurtosis are 9/5, 12/5, 3, and $3(v+w+1)\,[\,vw(v+w-6)+2(v+w)^2\,]\,/\,[\,vw(v+w+2)(v+w+3)\,]$.

Correlation of Normal(100,36) with Uniform(18,38)
Average computed skewness values, runs of size 25

	Normal(100,36)			Uniform(18,38)		
Theoretical correlation	10	100	500	10	100	500
-0.6	0.0304	-0.0150	0.0006	0.0159	0.0000	0.0000
-0.4	-0.0952	-0.0129	0.0031	-0.0174	0.0000	0.0000
-0.2	-0.0100	0.0232	-0.0006	-0.0061	0.0000	0.0000
0.3	0.1387	-0.0023	0.0055	0.0173	0.0001	0.0000
0.5	-0.0512	0.0206	-0.0008	-0.0059	0.0000	0.0000
0.7	-0.0764	0.0097	-0.0013	0.0000	0.0000	0.0000

Expected skewness values for N(100,36) and U(18,38) are 0 and 0 exactly

Correlation of Triangular(10,15,40) with Triangular(10,30,40)
Average computed skewness values, runs of size 25

	Triangular(10,15,40)			Triangular(10,30,40)		
Theoretical correlation	10	100	500	10	100	500
-0.6	0.5069	0.5036	0.5047	-0.2691	-0.3066	-0.3054
-0.4	0.4198	0.5048	0.5049	-0.3339	-0.3071	-0.3053
-0.2	0.4470	0.5068	0.5047	-0.2897	-0.3033	-0.3055
0.3	0.5394	0.5041	0.5049	-0.2095	-0.3043	-0.3052
0.5	0.4408	0.5062	0.5048	-0.3142	-0.3023	-0.3055
0.7	0.4319	0.5050	0.5047	-0.3299	-0.3047	-0.3055

Expected skewness values of T(10,15,40) and T(10,30,40) are 0.5047 and -0.3054 (4 d.p.s)

Correlation of Beta(12,15,30) with Beta(12,25,30)
Average computed skewness values, runs of size 25

	Beta(12,15,30)			Beta(12,25,30)		
Theoretical correlation	10	100	500	10	100	500
-0.6	0.5527	0.6491	0.6742	-0.3468	-0.4389	-0.4196
-0.4	0.5054	0.6731	0.6683	-0.4104	-0.4223	-0.4220
-0.2	0.5915	0.6546	0.6779	-0.3083	-0.4361	-0.4210
0.3	0.5347	0.6693	0.6707	-0.3751	-0.4136	-0.4224
0.5	0.4615	0.7052	0.6644	-0.4218	-0.4213	-0.4200
0.7	0.5222	0.6720	0.6717	-0.3860	-0.4428	-0.4197

Expected skewness values for B(12,15,30) and B(12,25,30) are 0.6742 and -0.4232 (4 d.p.s)

Correlation of Normal(100,36) with Uniform(18,38)
Average computed kurtosis values, runs of size 25

	Normal(100,36)			Uniform(18,38)		
Theoretical correlation	10	100	500	10	100	500
-0.6	2.5513	2.9618	2.9788	1.8148	1.8000	1.8000
-0.4	2.4458	3.0523	2.9875	1.7907	1.8000	1.8000
-0.2	2.3752	2.9770	2.9827	1.7796	1.8000	1.8000
0.3	2.4850	2.9091	2.9796	1.7997	1.8001	1.8000
0.5	2.5223	2.9078	3.0038	1.7884	1.8000	1.8000
0.7	2.4412	2.9526	2.9887	1.7935	1.8000	1.8000

Expected kurtosis values for N(100,36) and U(18,38) are 3 and 1.8 exactly

Correlation of Triangular(10,15,40) with Triangular(10,30,40)
Average computed kurtosis values, runs of size 25

	Triangular(10,15,40)			Triangular(10,30,40)		
Theoretical correlation	10	100	500	10	100	500
-0.6	2.3263	2.3966	2.3999	2.3134	2.4013	2.3998
-0.4	2.1786	2.4025	2.4005	2.2833	2.4065	2.3998
-0.2	2.1921	2.4059	2.4000	2.2247	2.3979	2.4001
0.3	2.3712	2.3952	2.4005	2.2161	2.3940	2.3995
0.5	2.2103	2.4018	2.4005	2.3007	2.3922	2.4006
0.7	2.1681	2.4004	2.4000	2.3104	2.3978	2.4002

Expected kurtosis value for T(10,15,40)and T(10,30,40) are 2.4 and 2.4 exactly

Correlation of Beta(12,15,30) with Beta(12,25,30)
Average computed kurtosis values, runs of size 25

	Beta(12,15,30)			Beta(12,25,30)		
Theoretical correlation	10	100	500	10	100	500
-0.6	2.6606	3.1570	3.2957	2.4422	2.8105	2.7400
-0.4	2.6131	3.2737	3.2616	2.4836	2.7477	2.7542
-0.2	2.7392	3.1835	3.3167	2.3181	2.8032	2.7468
0.3	2.6138	3.2604	3.2745	2.4457	2.7210	2.7524
0.5	2.4587	3.4315	3.2386	2.4471	2.7488	2.7430
0.7	2.6509	3.2816	3.2810	2.4802	2.8230	2.7402

Expected kurtosis values for B(12,15,30) and B(12,25,30) are 3.2888 and 2.7548 (4 d.p.s)

Appendix VIII

Further analysis of the sample product moment correlation coefficient between two variables.

The results complete the cases referred to in section 6.3.2 and 6.3.3.

Combining U[0,1) with U[0,1) via the RCM: Case (2): $\rho = -0.4$

Distribution of the Generated Correlation Coefficients

Sample Size:	10	100	500
Mean	-0.3994	-0.4000	-0.4000
Median	-0.4000	-0.4000	-0.4000
Std.Dev	0.0031	0.0000	0.0000
Range	0.0129	0.0000	0.0000
Minimum	-0.4064	-0.4000	-0.4000
Maximum	-0.3935	-0.4000	-0.4000

Combining U[0,1) with U[0,1) via the RCM: Case (3): $\rho = -0.2$

Distribution of the Generated Correlation Coefficients

Sample Size:	10	100	500
Mean	-0.2005	-0.2000	-0.2000
Median	-0.2005	-0.2000	-0.2000
Std.Dev	0.0037	0.0000	0.0000
Range	0.0198	0.0000	0.0000
Minimum	-0.2123	-0.2000	-0.2000
Maximum	-0.1925	-0.2000	-0.2000

Combining U[0,1) with U[0,1) via the RCM: Case (4): $\rho = 0.3$			
Distribution of the Generated Correlation Coefficients			
Sample Size:	10	100	500
Mean	0.2989	0.3000	0.3000
Median	0.2992	0.3000	0.3000
Std.Dev	0.0035	0.0000	0.0000
Range	0.0154	0.0000	0.0000
Minimum	0.2919	0.3000	0.3000
Maximum	0.3073	0.3000	0.3000

Combining U[0,1) with U[0,1) via the RCM: Case (5): $\rho = -0.5$			
Distribution of the Generated Correlation Coefficients			
Sample Size:	10	100	500
Mean	0.5001	0.5000	0.5000
Median	0.5000	0.5000	0.5000
Std.Dev	0.0051	0.0000	0.0000
Range	0.0259	0.0000	0.0000
Minimum	0.4866	0.5000	0.5000
Maximum	0.5126	0.5000	0.5000

Combining U[0,1) with U[0,1) via the RCM: Case (6): $\rho = 0.7$

Distribution of the Generated Correlation Coefficients

Sample Size:	10	100	500
Mean	0.7000	0.7000	0.7000
Median	0.7003	0.7000	0.7000
Std.Dev	0.0026	0.0000	0.0000
Range	0.0119	0.0000	0.0000
Minimum	0.6941	0.7000	0.7000
Maximum	0.7060	0.7000	0.7000

Combining T(10,30,40) with B(12,15,30) via the RCM:
Case (2): $\rho = -0.4$

Distribution of the Generated Correlation Coefficients

	Using the Correlation Model *Before* Swapping			Using the Correlation Model *After* Swapping		
Sample Size:	10	100	500	10	100	500
Mean	-0.3953	-0.4180	-0.4173	-0.3998	-0.4000	-0.4000
Median	-0.4154	-0.4179	-0.4172	-0.4000	-0.4000	-0.4000
Std.Dev.	0.0806	0.0116	0.0030	0.0021	0.0000	0.0000
Range	0.3412	0.0402	0.0128	0.0090	0.0000	0.0000
Minimum	-0.5500	-0.4381	-0.4249	-0.4043	-0.4000	-0.4000
Maximum	-0.2088	-0.3980	-0.4121	-0.3953	-0.4000	-0.4000

Combining T(10,30,40) with B(12,15,30) via the RCM:
Case (3): $\rho = -0.2$

Distribution of the Generated Correlation Coefficients

	Using the Correlation Model ***Before*** Swapping			Using the Correlation Model ***After*** Swapping		
Sample Size:	10	100	500	10	100	500
Mean	-0.2174	-0.2209	-0.2220	-0.2007	-0.2000	-0.2000
Median	-0.1900	-0.2183	-0.2215	-0.2005	-0.2000	-0.2000
Std.Dev	0.0795	0.0142	0.0035	0.0032	0.0000	0.0000
Range	0.2873	0.0618	0.0113	0.0173	0.0000	0.0000
Minimum	-0.3681	-0.2593	-0.2281	-0.2102	-0.2000	-0.2000
Maximum	-0.0808	-0.1975	-0.2168	-0.1929	-0.2000	-0.2000

Combining T(10,30,40) with B(12,15,30) via the RCM:
Case (4): $\rho = 0.3$

Distribution of the Generated Correlation Coefficients

	Using the Correlation Model ***Before*** Swapping			Using the Correlation Model ***After*** Swapping		
Sample Size:	10	100	500	10	100	500
Mean	0.2886	0.3228	0.3232	0.2999	0.3000	0.3000
Median	0.2863	0.3187	0.3235	0.2999	0.3000	0.3000
Std.Dev.	0.0581	0.0132	0.0026	0.0021	0.0000	0.0000
Range	0.2161	0.0486	0.0089	0.0073	0.0000	0.0000
Minimum	0.1858	0.3028	0.3187	0.2962	0.3000	0.3000
Maximum	0.4018	0.3513	0.3276	0.3035	0.3000	0.3000

Combining T(10,30,40) with B(12,15,30) via the RCM:
Case (5): $\rho = 0.5$

Distribution of the Generated Correlation Coefficients

	Using the Correlation Model *Before* Swapping			Using the Correlation Model *After* Swapping		
Sample Size:	10	100	500	10	100	500
Mean	0.4783	0.5048	0.5065	0.5002	0.5000	0.5000
Median	0.4762	0.5033	0.5063	0.5010	0.5000	0.5000
Std.Dev.	0.0585	0.0099	0.0024	0.0025	0.0000	0.0000
Range	0.2644	0.0387	0.0105	0.0098	0.0000	0.0000
Minimum	0.3649	0.4920	0.5017	0.4948	0.5000	0.5000
Maximum	0.6292	0.5306	0.5121	0.5046	0.5000	0.5000

Appendix IX

Comparing the result of running the RCM with corresponding runs of @Risk.

These results complete the remaining cases referred to in section 6.4.2. and 6.4.3.

Combining U[0,1) with U[0,1): Case (2): ρ = - 0.4

Distribution of the Generated Correlation Coefficients: @RISK versus the RCM

Sample Size:	Using @RISK 10	100	500	Using the RCM 10	100	500
Mean	-0.5337	-0.5293	-0.5357	-0.3994	-0.4000	-0.4000
Median	-0.5725	-0.5386	-0.5438	-0.4000	-0.4000	-0.4000
Std.Dev	0.1637	0.0650	0.0302	0.0031	0.0000	0.0000
Range	0.6397	0.2627	0.1120	0.0129	0.0000	0.0000
Minimum	-0.8305	-0.6503	-0.5951	-0.4064	-0.4000	-0.4000
Maximum	-0.1907	-0.3876	-0.4831	-0.3935	-0.4000	-0.4000

Combining U[0,1) with U[0,1): Case (3): ρ = - 0.2

Distribution of the Generated Correlation Coefficients: @RISK versus the RCM

Sample Size:	Using @RISK 10	100	500	Using the RCM 10	100	500
Mean	-0.0691	-0.2336	-0.2269	-0.2005	-0.2000	-0.2000
Median	-0.1375	-0.2162	-0.2258	-0.2005	-0.2000	-0.2000
Std.Dev	0.3890	0.1033	0.0329	0.0037	0.0000	0.0000
Range	1.5355	0.3501	0.1345	0.0198	0.0000	0.0000
Minimum	-0.7158	-0.4232	-0.2918	-0.2123	-0.2000	-0.2000
Maximum	0.8197	-0.0731	-0.1574	-0.1925	-0.2000	-0.2000

Combining U[0,1) with U[0,1): Case (4): ρ = 0.3

Distribution of the Generated Correlation Coefficients
@RISK versus the RCM

Sample Size:	Using @RISK 10	100	500	Using the RCM 10	100	500
Mean	0.3329	0.3696	0.3751	0.2989	0.3000	0.3000
Median	0.4403	0.3662	0.3848	0.2992	0.3000	0.3000
Std.Dev	0.2730	0.0771	0.0381	0.0035	0.0000	0.0000
Range	1.0408	0.3037	0.1347	0.0154	0.0000	0.0000
Minimum	-0.3360	0.1906	0.3012	0.2919	0.3000	0.3000
Maximum	0.7048	0.4943	0.4359	0.3073	0.3000	0.3000

Combining U[0,1) with U[0,1): Case (5): ρ = 0.5

Distribution of the Generated Correlation Coefficients
@RISK versus the RCM

Sample Size:	Using @RISK 10	100	500	Using the RCM 10	100	500
Mean	0.6743	0.7054	0.6966	0.5001	0.5000	0.5000
Median	0.7079	0.7092	0.6957	0.5000	0.5000	0.5000
Std.Dev	0.1534	0.0413	0.0111	0.0051	0.0000	0.0000
Range	0.6431	0.1497	0.0499	0.0260	0.0000	0.0000
Minimum	0.2519	0.6243	0.6720	0.4866	0.5000	0.5000
Maximum	0.8949	0.7740	0.7219	0.5126	0.5000	0.5000

Combining U[0,1) with U[0,1): Case (6): $\rho = 0.7$

Distribution of the Generated Correlation Coefficients @RISK versus the RCM

Sample Size:	Using @RISK 10	100	500	Using the RCM 10	100	500
Mean	0.9293	0.9244	0.9243	0.7000	0.7000	0.7000
Median	0.9260	0.9245	0.9244	0.7003	0.7000	0.7000
Std.Dev	0.0206	0.0058	0.0021	0.0026	0.0000	0.0000
Range	0.0706	0.0217	0.0085	0.0119	0.0000	0.0000
Minimum	0.9008	0.9135	0.9199	0.6941	0.7000	0.7000
Maximum	0.9714	0.9352	0.9284	0.7060	0.7000	0.7000

Combining T(10,30,40) with B(12,15,30): Case (2): $\rho = -0.4$

Distribution of the Generated Correlation Coefficients: @RISK versus the RCM

Sample Size:	Using @RISK 10	100	500	Using the RCM **After** Swapping 10	100	500
Mean	-0.3692	-0.3870	-0.3899	-0.3998	-0.4000	-0.4000
Median	-0.3932	-0.4039	-0.3906	-0.4000	-0.4000	-0.4000
Std.Dev.	0.2604	0.0679	0.0437	0.0021	0.0000	0.0000
Range	1.0857	0.2170	0.1820	0.0090	0.0000	0.0000
Minimum	-0.7316	-0.4725	-0.4702	-0.4043	-0.4000	-0.4000
Maximum	0.3541	-0.2555	-0.2882	-0.3953	-0.4000	-0.4000

Combining T(10,30,40) with B(12,15,30): Case (3): ρ = -0.2

Distribution of the Generated Correlation Coefficients: @RISK versus the RCM

Sample Size:	Using @RISK 10	100	500	Using the RCM **After** Swapping 10	100	500
Mean	-0.1080	-0.2070	-0.1974	-0.2007	-0.2000	-0.2000
Median	-0.1972	-0.2259	-0.2006	-0.2005	-0.2000	-0.2000
Std.Dev	0.3828	0.1118	0.0401	0.0032	0.0000	0.0000
Range	1.2722	0.4036	0.1782	0.0173	0.0000	0.0000
Minimum	-0.7077	-0.3611	-0.2751	-0.2102	-0.2000	-0.2000
Maximum	0.5645	0.0424	-0.0969	-0.1929	-0.2000	-0.2000

Combining T(10,30,40) with B(12,15,30): Case (4): ρ = 0.3

Distribution of the Generated Correlation Coefficients: @RISK versus the RCM

Sample Size:	Using @RISK 10	100	500	Using the RCM **After** Swapping 10	100	500
Mean	0.1775	0.2951	0.2990	0.2999	0.3000	0.3000
Median	0.1900	0.2849	0.3023	0.2999	0.3000	0.3000
Std.Dev.	0.3275	0.0714	0.0362	0.0021	0.0000	0.0000
Range	1.4309	0.2361	0.1544	0.0073	0.0000	0.0000
Minimum	-0.6646	0.1809	0.2128	0.2962	0.3000	0.3000
Maximum	0.7663	0.4170	0.3672	0.3035	0.3000	0.3000

Combining T(10,30,40) with B(12,15,30): Case (5): $\rho = 0.5$

**Distribution of the Generated Correlation Coefficients:
@RISK versus the RCM**

	Using @RISK			Using the RCM **After** Swapping		
Sample Size:	10	100	500	10	100	500
Mean	0.4507	0.4828	0.4855	0.5002	0.5000	0.5000
Median	0.4977	0.4952	0.4875	0.5010	0.5000	0.5000
Std.Dev.	0.2468	0.0682	0.0274	0.0025	0.0000	0.0000
Range	1.1927	0.2757	0.1166	0.0098	0.0000	0.0000
Minimum	-0.4426	0.3259	0.4235	0.4948	0.5000	0.5000
Maximum	0.7501	0.6016	0.5401	0.5046	0.5000	0.5000

Bibliography

Abhi, B. (1998). Practical risk analysis in scheduling. *Transactions-of-AACE-International*, pp.R1-R4.

Amirkhalili, R. (1997). Risk and capital budgeting. *Transactions of AACE International*. pp. 80-83.

Anders, J. (1994). Incorporating risk in exploration economics. *World Oil*, Sept, v215n8, pp.49-55.

Artikis, P.G. (1999). A capital budgeting stochastic simulation model applied in the banking industry. *Management Finance*, v25n8, pp.1-8.

Atkinson, S., Kelliher, C. and LeBruto, S. (1997). Capital budgeting decisions using "crystal ball". *Cornell Hotel and Restaurant Administration* Quarterly, Oct, v38n5, pp. 20-27.

Avramidis, A.N. and Wilson, J.R. (1996). Integrated variance reduction strategies for simulation. *Operations Research*, v44n2, March-April, pp.327-591.

Baker, M (1999). Discussion of sorting algorithms. http://atschool.eduweb.co.uk/mbaker/sorts.html

Baker, R.A. (1988). When is a prospect or play played out? *The Oil and Gas Journal*, Jan, pp. 77-80.

Baker, S., Ponniah, D., and Smith, S. (1998). Techniques for the analysis of risks in major projects. *Journal of Operational Research Society*, 49, pp. 567-572.

Balson, W.E., Welsh, J.L. and Wilson, D.S. (1992). Using decision anslysis and risk analysis to manage environmental risks. *Interfaces*. v22n6 pp. 126-139.

Bascomb, N. (1993). The risk analysts analyzed, *Euromoney*, Sept, n293, pp. 369-370.

Bazerman, M.H. (1994). *Judgement in managerial decision making*. Wiley, 3rd ed.

Behrenbruch, P., Azinger, K.L., Foley, M.V. (1989). Uncertainty and risk in petroleum exploration and development: *The expectation curve method*. Paper SPE 19475 presented at the SPR Asia-Pacific Conference held in Sydney.

Bennett, H,J, Thompson, J.G., Quiring, H.J. and Toland, J.E.(1970). *Financial evaluation of mineral deposits using sensitivity and probabilistic analysis methods*. Information Circular 8495, U.S. Dept of the Interior, Bureau of Mines.

Bierman, H (1986). *Quantitve analysis for business decisions.* Irwin. 7th ed.

Birnic, J. and Yates, A. (1991). Cost prediction using decision/risk analysis methodologies. *Construction Management and Economics*, v9n2, pp.171-186.

Bodily, S.E. (1992). Introduction: the practice of decision and risk analysis. *Interfaces*, 22:6, Nov-Dec, pp.1-4.

Brod, D. (1992). Quantifying uncertainty: risk analysis for forecasting and strategic planning. *Government Finance Review*, June, v8n3, pp.25-29.

Brookfield, D. (1995). Risk and capital budgeting: Avoiding the pitfalls in using NPV when risk arises. *Management Decision*, v33n8, pp.56-59.

Buchan, D.H. (1994). Risk analysis - Some practical suggestions. Cost *Engineering*, Jan, v36n1, pp.29-34.

Burley, T.A. and O'Sullivan G. (1986). *Work out Operational Research.* Palgrave.

Byrne, M.P. (1992). Examine risk factors. *Transportation and Distribution*, Nov, v33n11, pp.59-62.

Camps, J.A. (1996). Simple steps help minimise costs, risks in project contracts (engineering and construction contracts). *Oil and Gas Journal*, v94n4, pp.32-36.

Cario, M.C., and Nelson, B.L. (1997). *Modeling and Generating Random Vectors with Arbitrary marginal Distributions and Correlation Matrix.* Working paper, Northwestern University

Carter, E.E. (1972). What are the risks in risk analysis? *Harvard Business Review*, July-August, pp.72-82.

Cassens, D.L., Gibson, H. and Friday, J.S. (1993). Modeling lumber manufacturing processes using Monte Carlo computer simulation, *Forest Products Journal*, Sept, v43n9, pp.41-48.

Cavinato, J. (1990). How to perform a risk analysis. *Chilton's Distribution*, April, v89n4, pp.74-76.

Celia, C.W. (1969). *Introduction to Numerical Analysis.* McGraw-Hill.

Chadwell, H.P., Goitein, B., Horvah, P. and Webster, A. (1996). Financial criteria, capital budgeting techniques, and risk analysis of manufacturing firms. *Journal of Applied Business Research,* Winter, v13n1, pp.95-104.

Chansa, C. and Mount-Campbell, C. A. (1991). Decision criteria in capital budgeting under uncertainties: implications for future research. *International Journal of Production Economics*, v23n1-3, pp.25-35.

Chen, H.-F. (2001). Initialization for NORTA Generation of Random Vectors with Specified Marginals and Correlations. *INFORMS,* 13.4

Chapman, C.B. and Ward, S.C. (1997). *Project risk management: Processes, techniques and insights.* Wiley.

Clark, C.S. and Kleiman, M.A. (1992). Tools of financial analysis: Risk, return, and incremental analysis. *Topics in Health Care Financing,* Fall, v19n1, pp.26-33.

Clemen, R.T. (1991). *Making Hard Decisions: An introduction to decision analysis.* PWS-Kent.

Clemen, R.T. and Reilly, T. (1999). Correlations and copulas for decision and risk analysis. *Management Science,* Feb, v45n2, pp.208- 224.

Coates, J.F. (1994). Confronting the risks to your business. *Research Technology Management,* Jan/Feb, v37n1, pp.13-14.

Coats, P.K. and Chesser, D.C. (1982). Coping with business risk through probabilistic financial statements. *Simulation*, June, pp.111-121.

Cochrane, R.E. (1992). Using @RISK to predict project costs. *American Association of Cost Engineers Transactions,* v1, pp.F.3.1- F.3.5.

Coleman, M.S., Lynford, L, and Titus, J.F. (1995). A newly integrated and dynamic approach to securitized real estate risk analysis and valuation. *Real Estate Finance,* Spring, v12n1, pp.68-80.

Cooper, D.F. and Chapman, C.B. (1987). *Risk analysis for large projects: Models, methods and cases.* Wiley.

Cowell, F.B. (1991). Distribution functions in risk analysis. *Journal of Operational Research Society*, 42(3), pp.267-269.

Cozzolino, J.M. (1979). A new method for risk analysis. *Sloan Management Review*, Spring, pp.53-65.

Curwin, J. and Slater, R. (2002). *Quantitative methods for business decisions*. Thomson, 5th ed.

Daly, J. (1993). The 30-minute risk analysis. *Computerworld*, Nov, v27n48, p.68.

Dattatreya, R. (1995). Choosing the right hedge. *Corporate Finance*, Risk Management Yearbook, pp.8-11.

Davies, G.G. and Whiteside, M.W. (1991). *An integrated approach to prospect evaluation*. Paper SPE 23157 at the Offshore Europe Conference in Aberdeen.

De'Ath, N.G. (1997). Risk analysis in international exploration: a qualitative approach. *World Oil*, Sept, v218n9, pp.63-67.

Dickson, G.C.A. (1991). *Risk Analysis*. Witherby, 2nd ed.

Dickman, J.E. (1997). Risk analysis for environmental remediation projects: A comprehensive approach, *Cost Engineering*, Aug, v39n8, pp.38-44.

Dingle, J. (1997). *Project management: Orientation for decision makers*. Arnold.

Drury, C. and Tayles, M. (1997). The misapplication of capital investment appraisal techniques. Management Decision, 35/2, pp.86-93.

Dugdale, D. (1991). Is there a 'correct' method of investment appraisal? *Management Accounting*, May, pp.46-50.

Edwards, A.L. (1976). *An introduction to linear regression and correlation*. W. H. Freeman.

Edwards, D. and Hamson, M. (2001). *GUIDE to mathematical modelling*. Palgrave.

Eldin, J.J. and Boudry, J.C. (1982). Planning risk analysis in offshore projects: a realistic approach. *American Association Cost Engins Transactions*, pp.C3.1-C3.6.

Erb, C.B., Harvey, C.R. and Viskanta, T.E. (1996). Political risk, economic risk, and financial risk. *Financial Analysts Journal*, Nov/Dec, v52n6, pp.29-46.

Eschenbach, T. G. (1992). Quick sensitivity analysis for small projects and feasibility studies. *American Association of Cost Engineers Transactions*, v2, pp.L.6.1-L.6.8.

Eschenbach, T.G. (1996). Risk management through sensitivity analysis. *Transactions of AACE International*, pp.DRM41 - DRM46.

Evans, D.A. and Langford, B.E. (1992). A model for socially responsible new technology decisions. *Journal of Applied Business Research*, Winter, v9nl, pp.77- 87.

Evan, J.R. and Olson, D.L. (1998). *Introduction to simulation and risk analysis*. Prentice Hall.

Evans, M., Hasting, N. and Peacock, B. (2000). *Statistical distributions.* Wiley, 3rd ed.

Everett, M.D. (1986). A simplified guide to capital investment risk analysis. *Planning Review*, July, pp.32-36.

Eynon, P.J. (1988). Avoid the seven deadly sins of strategic risk analysis. *Journal of Business Strategy*, Nov/Dev, pp.18-22.

Farragher, E.J., Kleiman, R.T. and Sahu, A.P. (1999). Current capital investment practices, *Engineering Economist*, v44n2, pp.137- 150.

Feldman, S.J. (1983). Industry analysis and investment decision-making under conditions of uncertainty. *Managerial and Decision Economics*, v4, Issue 3, Corporate Strategy, Sept, pp.193-207.

Ferranti, M. (1993). Crystal Ball 3.0 gazes into 1-2-3. *PC Week*, August, v10n32, pp.33-35.

Fisher, I.N. and Hall, G.R. (1969). Risk and corporate rates of return. *Quarterly Journal of Economics*, February, pp.79-92.

Fong, G. and Vasicek, O.A. (1997). A multidimensional framework for risk analysis. *Financial Analysts Journal*, Jul/Aug, v53n4, pp.51- 57.

Frosdick, S. (1997). The techniques of risk analysis are insufficient in themselves. *Disaster Prevention and Management*. v6n3, pp.165-177.

Gable, J. (1992). Net Present Value: A financial tool for complicated times. *Records Management Quarterly*, Jan, v26n1, pp.3-5, 18.

Gapenski, L.C. (1990). Using Monte Carlo simulation to make better capital investment decisions. *Hospital & Health Services Administration*, Summer, v35n2, pp.207-210.

Gapenski, L.C. (1992). Accuracy of investment risk model varies. *Healthcare Financial Management*, April, v46n4, pp.40-52.

Garigliano, J. (1997). True value. *Folio the Magaxzine for Magazine Mangement*, Mar 1, v26n3, pp.34-37.

Ghosh, S., and Henderson, S.G (2002) (Cornell University). Properties of the NORTA Method in Higher Dimensions. *Proceedings of the 2002 Winter-Simulation Conference.*

Giaccotto, C. (1984). A simplified approach to risk analysis. *Engineering Economist*, Summer, v29n4, pp.273-286.

Greer, W.R. (1974). Theory versus practice in risk analysis: An empirical study. *The Accounting Review*, July, pp.496-505.

Goodwin, P. and Wright, G (1991). *Decision analysis for management judgement.* Wiley.

Hall, W.K. (1975). Why risk analysis isn't working. *Long Range Planning*, 8, pp.25-29.

Hanke, J.E., Reitsch, A.G.. and Wichern, D.W. (2001). *Business forecasting*. Prentice Hall International.

Harris, D.J. (1979). Risk analysis and financial planning. *Management Accounting*, July-August, pp.24-27.

Harris, D.J. and Williams, D.G (1985). The use of risk analysis within British Rail. *Managerial and Decision Economics*, v6, Issue 4, pp.202-209.

Hertz, D.B. (1964). Risk analysis in capital investment. *Harvard Business Review*, September - October, pp.169-181.

Hertz, D.B. and Thomas, H. (1984). *Practical risk analysis: An approach through case histories.* Wiley.

Hess, S.W. and Quigley, H.A. (1963). Analysis of risk in investment using Monte Carlo techniques. *Chemical Engineering Progress Symposium Series*, 59(42), pp.55- 63.

Higgins, J.G (1993). Planning for risk and uncertainty in oil exploration. *Long Range Planning*, v26n1, pp.111- 122.

Hillier, F.S. (1962). The deviation of probabilistic information for the evaluation of risky investments. *Management Science*, pp.443- 457.

Hillier, F.S. and Lieberman, G.J. (2005). *Introduction to Operations Research*. McGraw-Hill, 8th ed.

Ho, S.S.M. and Pike, R.H. (1991). Risk analysis in capital budgeting contexts: simple or sophisticated? *Accounting and Business Research*, v21n83, pp.227-238.

Hodder, J.E. (1985). Pitfalls in evaluating risky projects. *Harvard Business Review*, Jan/Feb, pp.128- 135.

Hoel, P.G. (1984). *Introduction to Mathematical Statistics*. Wiley, 5th ed.

Hoskins, C.G. (1975). Theory versus practice in risk analysis: An empirical study: A comment. *The Accounting Review*, October, pp.835- 843.

Hosseini, J. (1986). Decision analysis and its application in the choice between two wildcat oil ventures. *Interfaces*, V16n2, March – April, pp.75-85.

House, W.C. (1966). The usefulness of sensitivity analysis in capital investment decisions. *Management Accounting*, February, pp.22- 29.

Hubbard, D.W. (1997). Risk vs. return. *Informationweek*, Jun 30, n637, p.105.

Hull, J.C.(1980). *The evaluation of risk in business investment*. Pergamon.

Hull, T.E. Dobell, A.R. (1962). Random number generators. *SIAM Review*, v4n3, July, pp.230-254.

Hurley, W.J. (1998). On the use of Martingales in Monte Carlo approaches to multi-period parameter uncertainty in capital investment risk analysis. *Engineering Economist*, Winter, v43n2, pp.169-182.

Iman, R.L., and Conover, W.J. (1980). Small Sample Sensitivity Analysis Techniques for Computer Models, with an Application to Risk Assessment. *Communications in Statistics: Theory and Methods*, A 9:pp. 1749-1 874.

Iman, R.L. and Conover, W.J. (1982). A distribution-free approach to inducing rank correlation among input variables. *Commun. Statist. Simula. Computa*, v11n2, pp.311- 334.

http://www.Indoinc.com (1999).

Jablonowski, M. (1996). Using probabilistic risk analysis to improve risk management. *Risk management*, March, v43n3, pp.23-28.

Johnson, D. (1997). The triangular distribution as a proxy for the beta distribution in risk analysis. *The Statistician*, v46n3, pp.387-398.

Johnson, D, (1998). The robustness of mean and variance approximations in risk analysis. *Journal of the Operational Research Society*, 49, pp.253- 262.

Johnson, D. (2002). Triangular approximations for continuous random variables in risk analysis. *Journal of the Operational Research Society*, v53, pp.457-467.

Johnson, R.E. (1984). Risk analysis: not a crystal ball. *Infosystems*, 11/84, p.64.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), pp.263- 291.

Karady, G. (1985). Risk analysis: A powerful planning tool for decision makers. *Today's Executive*, 91(1), pp.8-11.

Karsak, E.E. (1998). Measures of liquidity risk supplementing fuzzy discounted cash flow analysis. *Engineering Economist*, Summer, v43n4, pp.331- 344.

Kelly, D.G. (1994). *Introduction to probability.* Macmillan.

Klein, J.H., Powell, P.L. and Chapman, C.B. (1994). Project risk analysis based on prototype activities. *Journal of Operational Research Society*, 45(7), pp.749-757.

Kliem, R.L. and Ludin, I.S. (1997). *Reducing project risk.* Gower.

Lamont, M. (2002). http://linux.wku.edu/~lamonl/algor/sort/sort.html.

Law, A.M. (1991). Simulation model's level of detail determines effectiveness. *Industrial Engineering.* Oct, v23n10, pp.16-18.

Lea, R.B., Adams, S.J., Boykin, R.F. and Smieliauskas, W. (1992). Modeling of the audit risk assessment process at the assertion level within an account balance; Discussions; Reply. *Auditing A Journal of Practice and Theory*, v11(Supplement), pp.152-197.

Lerche, I. And Mackay, J.A. (1999). *Economic risk in hydrocarbon exploration.* Academic Press.

Levy, H. and Sarnat, M. (1994). *Capital investment and financial decisions*. Prentice Hall, 5th ed.

Lewis, P. and Orav, E. (1989). *Simulation methodology for statisticians, operations analysts, and engineers.* Wadsworth & Brooks.

Li, S.T., and Hammond, J.L. (1975). Generalisation of Pseudo-Random Numbers with Specified Univariate Distributions and Correlation Coefficients. *IEEE Trans. on Systems, Management and Cybernetics*, 5, pp 557-561.

Linnerooth-Bayer, J. and Wahlstrom, B. (1991). Applications of probabilistic risk assessmeet: The Selection of Appropriate Tools. *Risk Analysis*, 11(2), pp.239-248.

Little, I.M.D. and Mirrlees, J.A. (1974). *Project appraisal and planning for developing countries*. Gower.

Lorance, R.B. (1992). Contingency draw-down using risk analysis. *American-Association-of-Cost-Engineers-Transactions*, v1, pp.F.6.1.-F.6.6.

Lucey, T. (1968). *Investment appraisal: Evaluating risk and uncertainty.* PhD thesis, Birmingham University.

Lurie, P.M., and Goldberg, M.S. (1998). An Approximate Method for Sampling Correlated Random Variables from Partially-Specified Distributions. *Management Science* Vol 44.2, pp 203-218

Macaulay, L. (1995). *Human-computer interaction for software designers*. International Thomson Publishing.

MacFarlane, J. (1995). The use of simulation in property investment analysis. *Journal of Property Valuation & Investment*, v13n4, pp.25-38.

Mackenzie, B. (1989). *Economic Analysis for Mine Evaluation and Operating Decisions.* Course notes presented at Workshop Course 629/89, Australian Mineral Foundation Incorporated.

Markowitz, H., Xu, G.L. and Putman, B. (1996). Deflating research expectation. *Global Investor*, Spe, n95, pp.57-59.

Marsaglia, G. and Olkin, I. (1984). Generating Correlation Matrices. *SIAM Jrnl.Scientific and Statistical Computing*, 5, pp 470-475

Matloff, N. (1988). *Probability introduction with applications to engineering and computer science*. PWS-Kent.

McCarthy, E.D. (1994). Using spreadsheet simulations in financial planning. *Journal of Financial Planning*, July, v7n3, pp.111-115.

McCray, A.W. (1975). *Petroleum evaluation and economic decisions.* Prentice-Hall.

McKay, M, D., Conover, W.J. and Beckman, R.J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, v21n2, pp.239-245.

McKim, R.A. (1992). Risk management: Back to basics. *Cost Engineering*, Dec, v34nl, pp.7-12.

Megill, R.E. (1984). *An introduction to risk analysis.* PennWell.

Megill, R. E. (1988). *An introduction to exploration economics.* PennWell, 3rd ed.

Meimban, J.J., Morris, J.S., and Govett, R.L. (1992). The evaluation of Wood-Fired congeneration investments using Monte Carlo simulation. *Engineering Economist*, Winter, v37n2, pp.115-136.

Moore, P.G. (1983). *The business of risk.* Cambridge University Press.

Moran, S.S. (1994). Another's man treasure. *Oil and Gas Investor*, Jan, v14n1, pp.35-38.

Morgan, M.G. and Henrion, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis.* Cambridge University Press.

Moser, K. and Coffee, P. (1989). Good choice for risk analysis. *PC Week*, Dec, v6n49, pp.37-39.

Mott, J.R.A. and Tumay, K. (1992). Developing a strategy for justifying simulation. *Industrial Engineering*, Jul, v24n7, pp.38-42.

Mulholland, H. and Jones, R.C. (1977). *Fundamentals of statistics.* Butterworth.

Murtha, J.A. (1995). *Decisions involving uncertainty: An a RISK tutorial for the petroleum industry.* Palisade.

Murtha, J.A. and Janusz, G.J. (1995). Spreadsheet generate reservoir uncertainty distributions. *The Oil and Gas Journal*, March, v93n11, pp.87-92

Myers, D.C. and Yeh, A.B. (1999). Generating correlated random variables for a simulation model. *Journal of Operational Research Society*, 50, pp.183-186.

Nanda, S. and Miller, A. (1996). Risk analysis in the valuation of medical practices. *Health Care Management Review*, Fall, v21n4, pp.26-32.

Naylor, T.H. and Thomas, C. (1983). Microeconomic foundations of corporate strategy. *Managerial and Decision Economics*, v4, Issue 3, Corporate Strategy, September, pp.127-135.

Newendorp, P.D. (1975). *Decision analysis for petroleum exploration.* Petroleum Publishing Company.

Newendorp, P.D. and Root, P.J. (1976). A method for treating dependencies between variables in simulation risk analysis model. *Journal of Petroleum Technology*, October, pp.1145-1150.

Newendorp, P.D. (1985). *Petroleum exploration economic and risk analysis: An overview of the analysis of profitability and risks of exploratory and development well drilling prospects.* Course Notes No. 30, Joint Association for Petroleum Exploration Courses (JAPEC, UK).

Olkin, I., Gleser, L.J. and Derman, C. (1994). *Probability models and applications.* Prentice Hall, 2nd ed.

Oldcorn, R. and Parker, D. (1995). *Strategic investment decision: evaluating opportunities in dynamic market.* Pitman.

Orczky, J. and Hancher, D. (1987). Schedule risk analysis using computer simulation. *American Association Cost Engins Transactions*, pp.A7.1-A7.6.

Palisade Corporation (1996). *Guide to using @RISK: Risk analysis and simulation add-in for Microsoft Excel or Lotus 1-2-3 window version.* Palisade.

Partowidagdo, W. and Cockcroft, P.J. (1989). *Petroleum investment in the Asia-Pacific region.* Paper SPE 19473 at the SPE Asia-Pacific Conference in Sydney.

Pearson, E.S. and Tukey, J.W. (1965). Approximate means and standard deviations based on distances between percentage points of frequency curves. *Biometrika*, v52, Issue 3/4, December, pp.533-546.

Phoa, W. (1999). Conditional Monte Carlo simulation. *Journal of Investing*, Fall, v8n3, pp.80-88.

Pidd, M. (1998). *Computer simulation in management science.* Wiley, 4th ed.

Pike, R. (1979). No accounting for intuition. *Management Accounting*, July/August, pp.20-21.

Pike, R. and Sharp, J. (1989). Trends in the use of management science techniques in capital budgeting. *Managerial and Decision Economics*, V10, Issue 2, June, pp.135-140.

Pouliquen, L.Y. (1970). *Risk analysis in project appraisal.* World Bank Staff Occasional Papers No. 11, International Bank for Reconstruction and Development.

Powell, S.G. (1991). The risk analysis of oil development in the Arctic National Wildlife Refuge. *Energy Journal*, v12n3, pp.55-76.

Price, J. (2002). An Implementation of the Lurie-Goldberg Algorithm in Schedule Risk Analysis. *Presented to the SCEA 2002 National Conference.* www.sceaOnline.net

Pryor, A.J. and Napier, J. (1985). The problems of medium and long term planning methods used by the oil industry. *Petroleum Review*, October.

Putnam, B. (1999). Visualizing risk and correlation. *Global Investor*, Jul/Aug, n124, pp.42-43.

Quality Assurance Review Guide (1999): http://www.dir.state.tx.us

Randhawa, S.U. and Douglas, J.A. (1993). Financial risk anlaysis using Financial Risk Simulation Program. *Industrial Management*, Sep/Oct, v35n5, pp.24-25.

Rappaport, A. (1967). Sensitivity analysis in decision making. *The Accounting Review*, July, pp.441- 456.

Richards, P. and Contesse, A. (1975). Risk analysis in the north sea. *Investment Analyst*, no.41, pp.20-25.

Ridley, T. (1999). Assessing risks and uncertainties in project development. *Petroleum Review*, February, pp.13-15 continued on p49.

Ripley, B.D. (1987). *Stochastic simulation.* Wiley.

Ritchie, B. and Marshall, D. (1993). *Business risk management.* Chapman & Hall.

Rosenhead, J.V. (1989). *Rational analysis for a problematic world: problem structuring methods for complexity, uncertainty and conflict*. *Wiley*.

Ross, S.A., Westerfield, R.W. and Jaffe, J. (1996). *Corporate finance*. Irwin, 4th ed.

Rowe, D.M. (1998). Value-at-risk: the correlated aggregation approach. *The Journal of Lending & Credit Risk Management*, v81 Issue 9, May, pp.51-53.

Rowe, W.D. (1977). *An anatomy of risk*. Wiley.

Sartori, D.E. and Smith, A.E. (1997). A metamodel approach to sensitivity analysis of capital project valuation. *Engineering Economist*, Fall, v43nl, pp.1-25.

Schmeiser, B. (1999). Advanced Input Modelling for Simulation Experimentation. *Proceedings of the 1999 Winter Simulation Conference*, Purdue University

Schuyler, J.R. (1992). Credible projections now require decision analysis. *Cost Engineering*, Mar, v34n3, pp.15-19.

Schuyler, J.R. (1999). Probabilistic reserves definitions, practices need further refinement. *The Oil and Gas Journal*, May 31, v97n22, pp.64- 67.

Sculli, D. (1981). The use of non-random numbers in Monte Carlo simulation. *The Statistician*, v30n2, June, pp.143-151.

Sheel, A. (1995). Monte Carlo simulations and scenario analysis - Decision-making tools for hoteliers. *Cornell Hotel and Restaurant Administration Quarterly*, Oct, v3n5, pp.18-26.

Silbergh, M. (1972). Profitability analysis - where are we now? *Journal of Petroleum Technology*, Jan, pp.90-100.

Singhvi, S.S. (1980). A practical approach to risk and sensitivity analyses. *Long Range Planning*, 13, pp.12-19.

Smith, D.J. (1994). Incorporating risk into capital budgeting decisions using simulation. *Management Decision*, v32n9, pp.20-26.

Smith, M.B. (1970). Probability models for petroleum investment decisions. *Journal of Petroleum Technology*, May, pp.543-550.

Smith, M.B. (1994). Understanding common risk analysis problems leads to better E&P Decisions. *World Oil*, 215(10), pp.95-97.

Stanley, L. (1992). Risk analysis: Using Lotus 1-2-3 with @RISK Add-In. *Journal of Cash Management*, Jan/Feb, v12nl, pp.45-48.

Sterk, W. and Vandenberg, P. (1992). A better way to evaluate hospital capital expenditures. *Healthcare Financial Management*, Jan, v46nl, pp.126- 127

Stuart, A., Ord, J. and Kendall Sir, M. (1994). *Kendall's advanced theory of statistics*. Arnold, London, 6th ed.

Swaney, J.A. (1996). Comparative risk analysis: Limitations and opportunities. *Journal of Economic Issues*, Jun, v30n2, pp.463- 473.

Taylor III, L.J. (1998). A comparison of capital budgeting models: locals versus global viewpoints. *Business Process Management Journal*, v4n4, pp.306-321.

Thompson, P. (1981). *Organization and economics of construction*. McGraw-Hill.

Tole, T.M., Hand, J.H. and McCord, S.O. (1997). The multinational firms' bias for risky capital projects. *Multinational Business Review*, Fall, v5n2, pp.10-15.

Trueman, R.E. (1977). *An introduction to quantitative methods for decision making*. Holt Rinehart and Winston.

Tuan, L.M. (1997). *Risk simulation analysis as a decision tool*. MSc dissertation, Imperial Collage.

Tversky, A. (1974). Assessing uncertainty. Journal of the Royal *Statistical Society*, 36(B), pp.148-159.

Uher, T.E. (1996). A probabilistic cost estimating model. *Cost Engineering*, 38(4), pp.33-40.

Uhler, R.S. and Bradley, P.G. (1970). A stochastic model for determining the economic prospects of petroleum exploration over large regions. *Journal of the American Statistical Association*, 65(330), pp.623-630.

Vale, C.D., and Maurelli, V.A. (1983). Simulating Multivariate Non-Normal Distributions. *Psychometrika*, 48,3, pp 465-471

Van Deventer, D.R. and Levin, J.W. (1996). Monte Carlo simulation? It's too risky. *American Banker*, Oct, v161n196, pp.18-19.

Van Dorp, J.R. and Duffey, M.R. (1999). Statistical dependence in risk analysis for project networks using Monte Carlo methods. *International Journal of Production Economics*, v58n1, pp.17-29.

Van Horne, J. (1995). *Financial management and policy.* Prentice Hall, 10th ed.

Van Rensburg, W.C.J. (1990). *Petroleum project economics and investment decision making.* Course notes presented at Workshop Course 672/90, Australia Mineral Foundation Inc.

Vose, D. (1996). *Quantitative risk analysis: A guide to Monte Carlo simulation modelling.* Wiley.

Vose, D. (2000). *Risk analysis: A quantitative guide.* Wiley, 2nd ed.

Wagle, B.A. (1978). Statistical Analysis of Risk in Capital Investment Projects. *Operational Ressarch Quarterly*, 18(1), pp.13-33.

Wagner, H.M. (1975). *Principles of operations research: with applications to managerial decisions.* Prentice Hall.

Wah, C.C. and Carole, F.J.H.C. (1993). A challenge to the investment appraisal of environmental protection projects. *Management Services*, May, v37n5, pp.12-15.

Wall, D.M. (1997). Distributions and correlations in Monte Carlo simulation. *Construction Management and Economics*, 15, pp.241- 258.

Wansbough, R.S. (1960). An approach to the evaluation of oil production capital investment risks. *Journal of Petroleum Technology*, Sept, pp.25-30.

Ward, C. (2002). *Managing project risk and uncertainty.* Wiley.

Watt Committee on Energy (1990). *Risk and the energy industries.* Paper presented at the 27th Consultative Conference, Birmingham.

Wendling, R.V. and Lorance, R.B. (2000). Intergration of schedule and cost risk models. *Cost-Engineering*, v42n4, pp.38-43.

White, D. (1995). Application of systems thinking to risk management. *Management Decision*, v33n10, pp.35-45.

William, T. (1995). A classified bibliography of recent research relating to project risk management. *European Journal of Operational Research*, v85n1, pp.18 -38

Wilson, J.R. (1984). Variance reduction techniques for digital simulation. *Journal of Mathematics and Management Science*, v4, pp.277-312.

Wilson, J.R. (1996). Integrated variance reduction strategies for simulation. *Operations Research*, Mar/Apr, v44n2, pp.327-346.

Wilson, J.R. (1998). Correlation-induction techniques for estimating quantiles in simulation experiments. *Operations Research*, Jul/Aug, v46n4, pp.574-591.

Woinsky, S.G. (1996). Economic analysis for the CPI. *Chemical Engineering*, July, pp.120-124.

Yeager, G. (1996). The best method for accurate risk management. *Corporate Finance*, Nov, n144, pp.32-34.

Zhi, H. (1993). Simulation analysis in project life cycle cost. *Cost Engineering*, Dec, v35n12, pp.13-17.